REMARKS

Applicant acknowledges and appreciates the Examiner's entrance of the Applicant's request for continued examination under 37 CFR 1.114 filed on December 16, 2002.

Interview Under 37 C.F.R. § 1.133, MPEP § 713.04

Applicant appreciates the personal interview granted by Examiner Fox on July 16, 2003 to Dr. Henry Daniell, the inventor, and his representative James Bauersmith in order to resolve the remaining issues in the case.

In light of the interview, the Applicant has amended the claims to resolve minor informalities and to further place the remainder of the claims in condition for allowance.

Non-Statutory Obvious-type Double Patenting Rejection

Applicant acknowledges with thanks the withdrawal of the Obvious-type Double Patenting Rejection over application Serial No. 08/972,901.

However, claims 3, 171, 190 – 192 are rejected under the judicially-created double-patenting doctrine over claims 19 – 23, 26 – 29, 31 and 34 of U.S. Patent No. 5,932,479. Applicant has addressed this rejection in a separate filing submitted on even date.

Indefiniteness – Rejections pursuant to 35 U.S.C. §112, 2d paragraph

Applicant acknowledges with appreciation the withdrawal of the indefiniteness rejections detailed in the previous office action. Likewise, acknowledges with appreciation the withdrawal of new matter rejections for claims 3, 171, 190, 191, 193 and 196 – 199.

Claims 193, 214 and 215 are rejected as being indefinite. By the foregoing amendments, Applicant has cured the indefiniteness of claim 190 by amending it to read "into a" in the penultimate line. Applicant has also deleted the reference to "which plant is the same as or different from the target higher plant" in claim 214, therefore curing the indefiniteness of that

claim and its dependant claim 215. Withdrawal of the rejections under 35 U.S.C. §112, second paragraph is respectfully requested.

Enablement Rejection – 35 U.S.C. §112, 1st paragraph

Claims 3, 171, 190 – 192, 196 – 199 remain rejected under 35 U.S.C. §112, first paragraph. Applicant respectfully traverse the rejections.

Applicant has amended the claims to further point out that the vector of the Applicant's invention is directed into a transcriptionally active intergenic spacer region of the chloroplast genome. Support for this amendment can be found on page 6, lines 11-15 and page 9, lines 31-35. Further, Sugita et al. (of record) demonstrated the presence of over 60 such spacer regions in plastid genomes.(Table II). This knowledge along with characterization of the plastid genome in a number of plants (Maier et al. 1995), readily allows one skilled in the art to practice the Applicant's claimed invention. *See* Maier, Rainer et al., "Complete Sequence of the Maize Chloroplast Genomes: Gene Content, Hotspots of Divergence and Fine Tuning of Genetic Information by Transcript Editing", J. Mol. Biol, 251: (1995), 614-626; discussion of the Maize Plastome, pg. 615, Col. 2, line 19 to pg. 618, Col. 1, line 13; *see also* Figure 1B, Length Comparison of Completely Sequenced Higher Plant Plastomes; Figure 2, Comparison Between Graminean Plants.

To further support the use of spacer regions for transgene expression, the Applicant has also included several examples of successful use of Applicant's invention. For example, in Hermann et al. (1999) (copy enclosed), the authors describe the targeting of transgenes into intergenic spacer region between *psbE* operon and a *petA* gene, which is known to be a suitable target site for the stable integration of transgenes. Hermann, Marita et al. "Transfer Of Plastid RNA-Editing Activity To Novel Sites Suggests A Critical Role For Spacing In Editing-Site

Recognition" Proc. Natl. Acad. Sci. USA 96 (1999): 4856-4861; *see* page 4858, lines 19 – 25. Another example is illustrated in Ruf et al. (2001) (copy enclosed), which used yet another spacer region of to transform tomato chloroplast. Ruf, Stephanie et al. "Stable Genetic Transformation Of Tomato Plastids And Expression Of A Foreign Protein In Fruit" Nature Biotechnology 19 (2001): 870 – 875. Ruf et al. 2001 illustrates the successful integration of transgenes between several tRNA Gly and tRNAfMet genes located between the psaB , psbC and psbD operon. *See id.* at page 874, column 2, lines 5 – 11. Furthermore, the Applicant has included, for the Examiner's convenience, a Table (citations omitted herein, but listed within the table) summarizing a multitude of spacer regions in which a number of foreign genes have been stably integrated into chloroplast genomes. As a result, the Applicant respectfully submits that these references support the use of any of a number of transcriptionally active spacer regions for the expression of transgenes and thus, one skilled in the art could predictably use any of a number of spacer regions to target transgenic expression.

Newly Amended Claims

Turning now to the Applicant's use of the phrase "either the 5' end or the 3' end, but not both" in the newly amended claims, Applicant respectfully submits that support for the use of such a phrase can be found in any of a number of examples described in the Applicant's specification, including Examples 2-16. Applicant submits that these transcriptionally active spacer regions are suitable for the insertion of transgenes without the need for 5' or 3' regulatory sequences, which can be expressed in the plastids of higher plants. One skilled in the art could readily use any identified spacer region to integrate foreign genes, without the use of regulatory sequence. Ruiz, Oscar et al. "Phytoremediation of Organomercurial Compounds Via Chloroplasts Genetic Engineering" Plant Physiol. 132 (2003): 1-9, *see* pg. 2, Col. 2, lines 24-29,

38-41. Applicant has shown several examples of stable integration and expression of foreign genes without using a 3' region or a promoter upstream of coding sequences, irrespective of the spacer region where foreign genes were integrated (rbcL/accD or trnI/trnA). In all of the expression cassettes described in the specification, the aadA gene has no 3' region and the gene of interest is without a promoter. These examples, which utilize regulatory elements of the plastid genome provide for the construction of a promoterless expression cassette wherein the gene coding for the peptide of interest can be driven by a native or an inserted promoter contained within the plastid genome but such a promoter is not present immediately upstream of the gene coding for the peptide of interest. The Applicant respectfully submits that a number of examples have illustrated that the gene coding for the peptide of interest does not contain a promoter directly upstream. Specifically, Figures 2A, 2B, 3A, 3B, 7B, 7D, 8, 25 and examples 1, 10, 11 and 16 show a gene of interest with no promoter directly upstream. The genes of interest, including genes conferring herbicide resistance, protein based polymer, insect resistance genes were expressed without a promoter. As another example, illustrating the use of such spacer regions with or without the need for 5' or 3' regulatory sequences, the Applicant has enclosed herewith a copy of Ruiz et al. (cited above) at page 169, Fig. 2B, wherein the pLDR-MerAB-3'-UTR and pLDR-MerAB vectors were constructed and successfully inserted into a spacer region of the plastid genome.

Prior Art Rejection – Anticipation

Claims 3, 171, 190 – 192 remains rejected over various prior art references. Claim 192 has been cancelled. Rejections against claims 3, 171, 190 – 191 remain. Applicant respectfully traverses.

As was discussed in the Examiner's interview, Applicant respectfully submit that Staub et al. (of record) does not teach the intergenic spacer region between the "universally" present *rbc*L and *acc*D chloroplast genes, because as is taught in Maier et al. (cited above), pg. 619, Figure 2B, the *rbc*L and *acc*D genes are not together in monocot chloroplast genomes. Specifically the *acc*D chloroplast gene does not exist in monocot chloroplast genomes, and is not universally present near *rbc*L in monocots.

The Authorities Relied Upon by the Examiner to Reject the Written Description of the Subject Invention Support the Applicant's Written Description of the Subject Invention

Applicant's attorney does not question the court's statement that *Universtiy of California v. Eli Lilly and Co.,* 119 F.3d 1559, 1568; 43 USPQ2d 1398, 1406 (Fed. Cir. 1997), "an invention requires precise definition, such as by structure, formula, [or] chemical name, of the claimed subject matter sufficient to distinguish it from other materials". In the instant application, the claims define the elements which comprise the universal integration and expression vector which comprises an expression cassette comprising a DNA sequence, both elements being defined for their function (coding, organization, etc.) The Examiner found but for some clarification in the claims, which are now corrected in this Amendment, that the claims defined the subject matter sufficiently to distinguish it from other material described in the prior art. Accordingly, the instant description is in accord with the *University of California* case.

The *MPEP* Section 2163, page 156 of Chapter 2100 of the August 2001 version, column 2, bottom paragraph, which the Examiner has relied upon, has been reviewed. Applicant's counsel submit that the DNA sequence is defined by its coding sequence for a peptide of interest, and the method of stable integration, whereby double homologous recombination, is facilitated.

Therefore, the instant disclosure is in accordance with MPEP as cited above and the Guidelines published in *Federal Register*/ Vol. 66, No. 4/Friday, January 5, 2001/ Notices: pp. 1099-1111.

Likewise, *Amgen Inc. v. Chugai Pharmaceutical Co. Ltd.*, 18 USPQ 2d 1016 at 1021, (Fed. Cir. 1991) is relied upon for the statement that a gene is not reduced to practice until the inventor can define it by "its physical or chemical properties" (e.g. a DNA sequence). *Amgen,* which deals with a purification of Erythropoietin (EPO) is not pertinent to the claims of this application. The "cassette" is well defined by its elements, properties and function. The *University of California* case (citation omitted) has been discussed above.

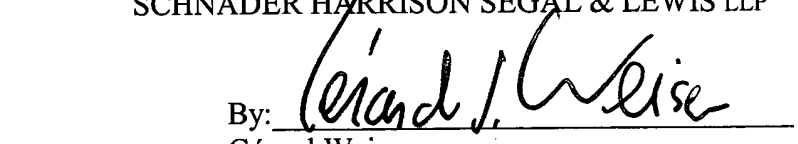
The Examiner relies on MPEP 2111.02, *Pitney Bowes, Inc. v. Hewlett-Packard Co.*, 51 USPQ2d 1161, 1165 (Fed. Cir. 1999), *Rowe v. Dror*, 42 USPQ2d 1550, 1553 (Fed. Cir. 1997), and *In re Schreiber*, 44 USPQ2d 1429, 1431 (Fed. Cir. 1997) for the proposition that an intended use is not given patentable weight in product claims. Applicant respectfully asserts that the claim preamble, in its previous form, was necessary to give life, meaning and vitality to the claim in that it taught the universal applicability of the vector. Nevertheless, the current claims have now been amended to expedite prosecution. Claim 192 has been cancelled; therefore the rejection is moot.

Applicant acknowledges the availability of later publications as prior art under the limited circumstances that the publication serves to illustrate a universal fact. MPEP 2124 and 2131.01, part (III). However, as discussed above, Staub et al. (1995) does not "teach a region comprising the intergenic spacer region between the universally present *rbcL* and *accD* chloroplast genes" because the *acc*D chloroplast gene does not exist in monocot chloroplast genomes, and is not universally present near *rbc*L in monocots.

Conclusion

In view of the forgoing, Applicant respectfully submits that the remainder of the claims are now in condition for allowance, which action is respectfully requested.

Respectfully submitted,
SCHNADER HARRISON SEGAL & LEWIS LLP

By: 

Gérard Weiser
Registration No. 19,783
1600 Market Street, Suite 3600
Philadelphia, Pennsylvania 19103
Telephone: 215-751-2811
Facsimile: 215-751-2658
E-mail: gweiser@schnader.com
Attorneys for Applicant

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Art Unit	: 1638	**22469**
Examiner	: D. T. Fox	
Serial No.	: 09/079,640	
Filed	: May 15, 1998	
Inventors	: Henry Daniell	Docket No. 1018-00
Title	: UNIVERSAL CHLOROPLAST	
	: INTEGRATION AND	
	: EXPRESSION VECTOR,	Confirmation No.: 8567
	: METHOD OF USE AND	
	: TRANSFORMED PLANTS	Dated: 09/11/2003

TRAVERSAL OF THE PROVISIONAL OBVIOUSNESS–TYPE DOUBLE PATENTING REJECTION

Claims 3, 171, 190 – 192 of the Application are rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 19 – 23, 25 – 29, 31 and 34 of U.S. Patent No. 5,932,479 ("the Patent"). The PTO states that "although the conflicting claims are not identical, they are not patentably distinct from each other." See PTO Office Actions of March 19, 2001, November 20, 2001, June 17, 2002 and March 11, 2003.

The double patenting rejection is respectfully traversed.

The Claims of the Application

The claims of the Application in issue are claims 3, 171 and 190 – 192. The broadest claim is claim 190, intermediate claims are claim 171 (transposon), claim 3 (nucleotide sequence coding for a selectable phenotype), claim 191 (selectable marker, promoter, control sequences . . . either of the 5' end or downstream from the 3' ends, but not both).

Applicant submits that the subject claims in the Application define an invention which is not "merely an obvious variation of the invention" claimed in the Patent claims identified above. The Applicant by his attorney shows that the claims in issue in the Application call for features which are unobvious over the claims of the Patent.

Claim 190 of the Application calls for a heterologous DNA sequence "coding for a peptide of interest." In addition, claim 190 calls for "control sequences positioned upstream from either of the 5' end or downstream of the 3' end, but not both." Also, the claim calls for a stable integration of the heterologous coding sequence "directed into a transcriptionally active intergenic spacer region of the chloroplast genome."

Claim 3 (claim 190), additionally calls for "a heterologous nucleotide sequence coding for a selectable phenotype."

Claim 171 (claim 190), additionally calls that the vector of claim 190 does not include a "transposon."

Claim 191 (an independent claim), in addition to the elements recited in claim 190, calls for "a selectable marker." Claim 191 also calls for "a promoter" which drives either one of a "selectable marker" or the "DNA sequence encoding a peptide of interest."

Claim 192 was cancelled without prejudice.

The Claims of the Patent

Claim 19 of the Patent calls for a "coding sequence." This claim is silent with respect to "a peptide of interest." Claim 19 is silent with respect to positioning a control sequence either upstream of the 5' end or downstream of the 3' end. The claim is silent with respect to directing the "integration of the heterologous coding sequence into a transcriptionally active spacer region of the chloroplast genome of the target plant."

Claim 20 (claim 19) calls for a promoter which is a "chloroplast promoter."

Claim 21 (claim 20) calls for types of promoters.

Claim 22 (claim 19) calls for a selectable marker which encodes a "selectable phenotype."

Claim 23 (claim 22) calls for the types of selectable markers.

Claim 25 (another independent claim) of the Patent is silent on the positioning a control sequence either upstream of the 5' end or downstream of the 3' end. The claim is silent with respect to directing the stable "integration of the heterologous coding sequence into a transcriptionally active spacer region of the chloroplast genome of the target plant."

The above-enumerated differences establish that each one of the claims of the Application in issue contain features which call for at least an unobvious distinction over the claims of the Patent. Applicants further submit that the unobvious distinctions of the Application's claims are not suggested by the cited claims of the '479 patent.

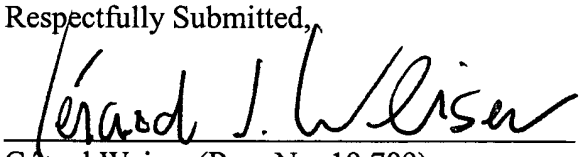
Conclusion

The claims 3, 171 and 190-191 of the Application are patentably distinct from the identified claims 19-23, 25-29, 31 and 34 of the Patent. The claims of the Application met the requirement of 35 U.S.C. 101 MPEP § 804, "One Way Obvious-Type." The claims of the Application meet the nonobviousness requirement of 35 U.S.C. 103.

Applicant respectfully requests the withdrawal of the non-statutory type double-patenting rejection of claims 3, 171, 190 – 191 of the Application. The claims are allowable, it is submitted.

In the event that the Examiner has any questions, regarding this document, he is invited to call the undersigned attorney at 215-751-2811.

Respectfully Submitted,



Gerard Weiser (Reg. No. 19,789)
Attorney for Applicant
Schnader Harrison Segal & Lewis, LLP
1600 Market Street, Suite 3600
Philadelphia, PA 19103
215-751-2811

RUIZ *ET AL.*

Phytoremediation of Organomercurial Compounds via Chloroplast Genetic Engineering[1]

Oscar N. Ruiz, Hussein S. Hussein, Norman Terry, and Henry Daniell*

Department of Molecular Biology and Microbiology, University of Central Florida, Orlando, Florida 32816–2360 (O.N.R., H.D.); and Department of Plant and Molecular Biology, University of California, Berkeley, California 94720–3102 (H.S.H., N.T.)

Mercury (Hg), especially in organic form is a highly toxic pollutant affecting plants, animals, and man. In plants, the primary target of Hg damage is the chloroplast; Hg inhibits electron transport and photosynthesis. In the present study, chloroplast genetic engineering is used for the first time to our knowledge to enhance the capacity of plants for phytoremediation. This was achieved by integrating a native operon containing the *merA* and *merB* genes (without any codon modification), which code for mercuric ion reductase (*mer*A) and organomercurial lyase (*mer*B), respectively, into the chloroplast genome in a single transformation event. Stable integration of the merAB operon into the chloroplast genome resulted in high levels of tolerance to the organomercurial compound, phenylmercuric acetate (PMA) when grown in soil containing up to 400 μM PMA; plant dry weights of the chloroplast transformed lines were significantly higher than those of wild type at 100, 200, and 400 μM PMA. That the merAB operon was stably integrated into the chloroplast genome was confirmed by polymerase chain reaction and Southern-blot analyses. Northern-blot analyses revealed stable transcripts that were independent of the presence or absence of a 3′-untranslated region downstream of the coding sequence. The merAB dicistron was the more abundant transcript, but less abundant monocistrons were also observed, showing that specific processing occurs between transgenes. The use of chloroplast transformation to enhance Hg phytoremediation is particularly beneficial because it prevents the escape of transgenes via pollen to related weeds or crops and there is no need for codon optimization to improve transgene expression. Chloroplast transformation may also have application to other metals that affect chloroplast function.

Mercury (Hg) pollution of soil and water is a world-wide problem (Dean et al., 1972; Krämer and Chardonnens, 2001). The extent to which Hg is harmful depends on the form of mercury present in the ecosystem. Inorganic forms of Hg are less harmful than organic forms partly because they bind strongly to the organic components of soil. For this reason, Hg does not tend to contaminate the ground water except when it leaches from a municipal landfill (U.S. Environmental Protection Agency, 1984). Organomercurial compounds, on the other hand, may be 200 times more toxic than inorganic Hg (Patra and Sharma, 2000) and methyl-Hg is especially toxic (Meagher and Rugh, 1997).

The principal forms of organomercurial compounds are alkyl mercurials (methyl- and ethyl-Hg), aryl mercurials (phenyl-Hg), and alkoxy alkyl Hg diuretics. The excessive use of organomercurial compounds (e.g. in fertilizers and pesticides) is known to have severe effects on plants. The main site of action of Hg damage appears to be the chloroplast thylakoid membranes and photosynthesis. Organomercurial compounds have been shown to strongly inhibit electron transport, oxygen evolution (Bernier et al., 1993), Hill reaction, photophosphorylation, and to quench chlorophyll fluorescence in photosystem II (Kupper et al., 1996). Furthermore, Prasad and Prasad (1987) showed that Hg might replace Mg from the chlorophyll moiety, leading to a reduction in chlorophyll content. Sen and Mondal (1987) and Sinha et al. (1996) reported a 26% (w/v) reduction of chlorophyll content in *Salvia natans* and 35% (w/v) in *Bacopa monnieri* at 5 μg mL^{-1} $HgCl_2$, even though these plants have a natural tolerance to Hg.

Current remediation methods to clean up heavy metal-contaminated soils include soil flushing, chemical reduction/oxidation and excavation, retrieval, and offsite disposal, all of which are expensive, environmentally invasive, and labor intensive (Kärenlampi et al., 2000). An alternative and more cost-effective approach is phytoremediation, i.e. the use of plants to clean up contaminated environments (Lin et al., 1995; Salt et al., 1995; Terry et al., 2000). With the aid of genetic engineering, plants can be genetically modified to substantially improve phytoremediation. Expression of several plant and bacterial genes in transgenic plants has significantly enhanced these plant remediation systems (Meagher, 2000; Doucleff and Terry, 2002). Several studies have successfully integrated bacterial genes into nuclear genomes to produce plants that were specifically engineered for phytoremediation of metal-polluted environments

[1] This work was supported in part by funding from Chlorogen Inc. (St. Louis)

* Corresponding author; e-mail daniell@mail.ucf.edu; fax 407–823–0956.

Article, publication date, and citation information can be found at www.plantphysiol.org/cgi/doi/10.1104/pp.103.020958.

(Heaton et al., 1998; Rugh et al., 1998; Nies, 1999). With respect to Hg, plants have been engineered with modified bacterial mercuric ion reductase (*mer*A) and organomercurial lyase (*mer*B) genes; these enzymes are capable of converting highly toxic methyl-Hg into the much less toxic Hg(0), which may then be volatilized (Rugh et al., 1996; Bizily et al., 1999, 2000).

All of the attempts to genetically engineer plants with improved phytoremediation have previously been based on transformation of the nuclear genome. An alternative and novel approach is to engineer the chloroplast genomes of higher plants. This approach offers several advantages over nuclear transformation, i.e. very high levels of transgene expression (up to 46% (w/w) of total protein; De Cosa et al., 2001), uniparental plastid gene inheritance (in most crop plants) that prevents pollen transmission of foreign DNA (Daniell et al., 1998; Daniell, 2002; Daniell and Parkinson, 2003), the absence of gene silencing (Lee et al., 2003) and positioning effect (Daniell et al., 2001a), the ability to express multiple genes in a single transformation event (De Cosa et al., 2001; Daniell and Dhingra, 2002), the ability to express bacterial genes without codon optimization (McBride et al., 1995; Kota et al., 1999; De Cosa et al., 2001), integration via a homologous recombination process that facilitates targeted transgene integration (Daniell et al., 2002), and sequestration of foreign proteins in the organelle, which prevents adverse interactions with the cytoplasmic environment (Daniell et al., 2001a; Lee et al., 2003). Engineering the chloroplast genome has successfully conferred insect resistance (McBride et al., 1995; Kota et al., 1999; De Cosa et al., 2001), herbicide resistance (Daniell et al., 1998), disease resistance (De Gray et al., 2001), drought tolerance (Lee et al., 2003), and expression of edible vaccines (Daniell et al., 2001a), monoclonals (Daniell, 2003), and biopharmaceuticals (Guda et al., 2000; Staub et al., 2000; De Gray et al., 2001; Fernandez-San Millan et al., 2003).

This is the first report where the chloroplast genome was engineered to enhance the capacity of plants for phytoremediation and where a native bacterial operon was used for expression in plants without codon optimization. Phenylmercuric acetate (PMA) was chosen to test the chloroplast transformation method because of the importance of toxicity of organomercurial compounds as environmental contaminants and because the site of action of organomercurial damage is the chloroplast (see above). The approach we used was to integrate a native operon containing the *mer*A and *mer*B genes, coding for mercuric ion reductase and organomercurial lyase, respectively, into tobacco (*Nicotiana tabacum*) chloroplast genomes. The results show that the chloroplast transgenic plants were substantially more resistant than wild type to the highly toxic organomercurial compound, PMA.

RESULTS AND DISCUSSION

Chloroplast Vectors and Bacterial Resistance Assays

The bacterial native genes, *mer*A (1.69 kb) and *mer*B (638 bp) that encode the mercuric ion reductase and the organomercurial lyase, respectively, were amplified by PCR from *Escherichia coli* strains harboring plasmids NR1 (containing the full-length *mer*A) and R831b (containing the full-length *mer*B). The PCR gene products were successively cloned into the pLD-vector, which is a chloroplast-specific vector used in previous publications from this laboratory (De Cosa et al., 2001; Daniell et al., 2001b). This vector contains the homologous recombination sequences (flanking sequences) that allow site-specific integration of the operon containing the *aadA*, *merB*, and *merA* genes into the inverted repeat region of the chloroplast genome in between the *trnI* (tRNA Ile) and *trnA* (tRNA Ala) genes (Daniell et al., 1998; Guda et al., 2000). The chloroplast 16S ribosomal RNA gene constitutive promoter (*Prrn*) drives the transcription of all downstream genes that include the *aad*A (aminoglycoside 3′-adenylyltransferase) gene conferring resistance to spectinomycin, the *merA*, and *merB* genes. Two versions of the chloroplast vector were made with the presence or absence of the 3′-untranslated region (UTR) from the chloroplast *psb*A gene that was expected to confer stability to transcripts, and they were designated pLDR-MerAB-3′-UTR and pLDR-MerAB, respectively (Fig. 1A). The pLDR-MerAB-3′-UTR and the pLDR-MerAB chloroplast vectors also contain the *E. coli* origin of replication and the ampicillin selectable marker that facilitates *E. coli* expression studies.

The transformed bacterial cells harboring pLDR-MerAB and pLDR-MerAB-3′-UTR, and the control untransformed cells (*E. coli*) were grown on Luria-Bertani medium in the presence of different concentrations of mercuric chloride. Bacterial cells containing the pLDR-MerAB and pLDR-MerAB-3′-UTR were able to grow in concentrations of $HgCl_2$ of up to 100 μM on solid agar plates (Fig. 1B). Untransformed *E. coli* cells were unable to grow even at a concentration of 25 μM. Although transformed cells were able to grow in liquid broth at concentrations of 25 and 50 μM $HgCl_2$, differences in the rate of growth between the clone transformed with the plasmid containing the 3′ terminator and the clone that lacked the terminator region were examined (Fig. 1C). It is known from previous studies that the 3′-UTRs in *E. coli* are engaged in the termination of transcription. The pLDR-MerAB-3′-UTR was expected to grow better in the presence of Hg because, by terminating effectively, more copies of a shorter transcript containing the *merAB* operon would be made, in contrast to fewer long transcripts in the case of the pLDR-MerAB clone. The Hg bioassay showed that indeed *E. coli* cells transformed with the pLDR-MerAB-3′-UTR vector resulted in higher bacterial growth when com-

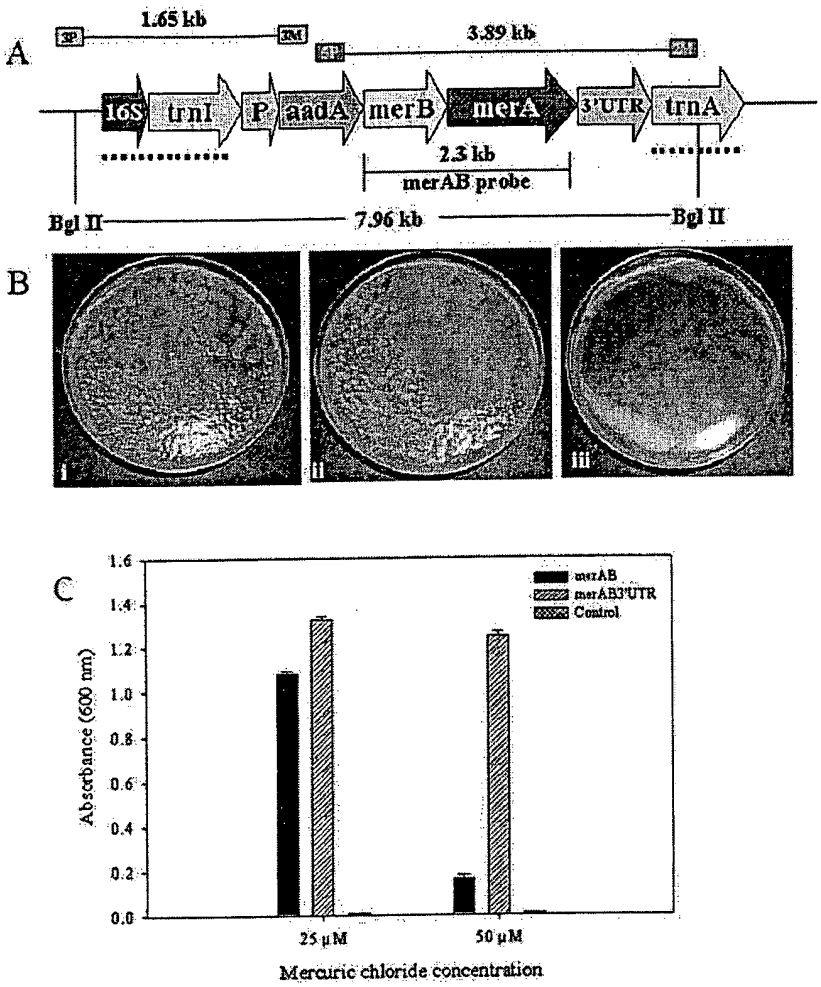


Figure 1. Bacterial bioassay. A, Schematic representation of the transformed chloroplast genome: The map shows the transgenic chloroplast genome containing the pLDR-MerAB-3′-UTR construct. The site-specific integration between *trn*I and *trn*A chloroplast genes is shown by the dotted line, specifying the homologous recombination sequences in the pLDR-MerAB-3′-UTR and pLDR-MerAB. Landing sites for the 3P/3M and 5P/2M primer pairs used in PCR confirmation of integration, and expected sizes of products are shown. *Bgl*II restriction digestion sites and the *mer*AB probe used in the Southern-blot analyses are shown. A fragment of 7.96 kb should be produced after restriction digestion of the transgenic chloroplast genome. B, Transformed *E. coli* grown in 100 μM $HgCl_2$. i, Transformed *E. coli* cells containing the vectors pLDR-MerAB; ii, pLDR-MerAB-3′-UTR grown in Luria-Bertani at 100 μM $HgCl_2$; iii, untransformed control (*E. coli*). C, Effect of mercuric chloride on *E. coli* cell proliferation. The transgenic clone pLDR-MerAB and pLDR-MerAB-3′-UTR and the control *E. coli* cells were grown on liquid Luria-Bertani medium with 25 and 50 μM of $HgCl_2$ for 24 h at 37°C. The A_{600} was measured.

pared with the bacterial cells containing the vector lacking a 3′ *psb*A-UTR (Fig. 1C).

Transformation, Selection, and Characterization of Chloroplast Transgenic Plants

Chloroplast-transgenic plants were obtained as described (Daniell, 1997). More than 20 positive independent transgenic lines were obtained with each construct. In this report, we show the results of two transgenic lines that were transformed with the pLDR-MerAB vector and the pLDR-MerAB-3′-UTR, respectively. The variability in expression levels among independent chloroplast-transgenic lines were minimal, as reported previously (Daniell et al., 2001a), and the results shown here correlate well with the results of other transgenic lines with the same chloroplast vectors.

The primer pair 3P and 3M was used to test integration of the transgene cassette into the chloroplast genome at very early stages during the selection process. The 3P primer lands in the native chloroplast genome and the 3M primer lands in the *aad*A gene that is present within the gene cassette (Fig. 1A). If integration has occurred, a 1.65 kb PCR product should be obtained (Fig. 2A). The untransformed control and the mutants (caused by the spontaneous mutation of the 16S rRNA gene that confers resistance to spectinomycin) did not show any product, confirming that these plants are negative for integration of transgenes (Fig. 2A). The integration of transgenes (*aad*A, *mer*A, and *mer*B) was further tested by using the 5P/2M primers and PCR analysis. The 5P and 2M primers annealed to the internal region of the *aad*A and *trn*A genes, respectively (Fig. 1A). The product size of positive transgenic clones was 3.89 kb, whereas the mutants and untransformed control did not show any PCR product (Fig. 2B). The DNA from full-grown T_0 and T_1 generation plants was extracted and used for the Southern-blot analysis (Fig. 3). The 0.81 kb flanking sequence probe that hybridizes with the *trn*I and *trn*A genes (Fig. 3A) allowed detection of the site-specific integration of the gene cassette into the chloroplast genome. The transformed chloroplast genome digested with *Bgl*II restriction enzyme produced a fragment of 7.96 kb

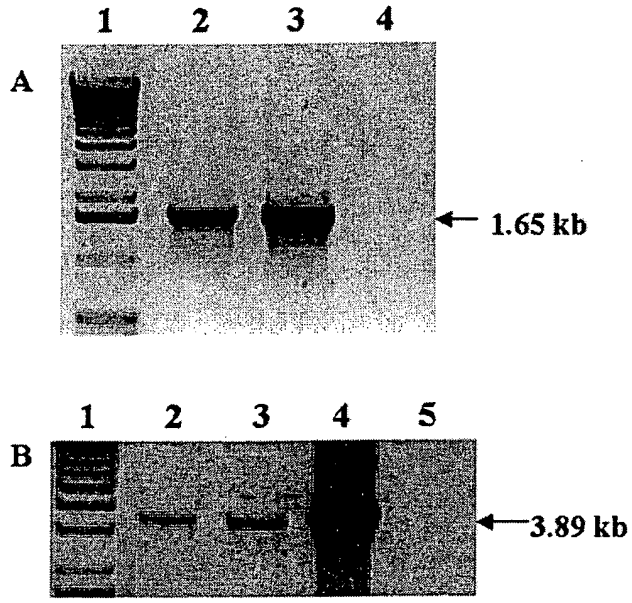


Figure 2. PCR analysis of control and putative transformants. A, PCR products (1.65 kb) using 3P/3M primers show integration into the chloroplast genome. Lane 1, Marker; lane 2, pLDR-MerAB transgenic line; lane 3, pLDR-MerAB-3′-UTR transgenic line; lane 4, untransformed wild type. B, PCR products (3.8 kb) using 5P/2M primers confirm *mer*AB integration. Lane 1, Marker; lane 2, pLDR-MerAB transgenic line; lane 3, pLDR-MerAB-3′-UTR transgenic line; lane 4, positive control (pLDR-MerAB plasmid DNA); lane 5, untransformed wild-type tobacco.

(Figs. 1A, and 3, B and C). The untransformed chloroplast genome digested with *Bgl*II yielded a 4.47-kb fragment (Fig. 3, A–C).

The flanking sequence probe also showed that homoplasmy of the chloroplast genomes was achieved through the selection process. Southern blots confirmed stable integration of foreign genes into all of the chloroplast genomes confirming homoplasmy. T_0 and T_1 generation transgenic plants only showed a single fragment of 7.96 kb. The absence of any detectable native untransformed chloroplast genomes not only confirmed homoplasmy, but also facilitated detection of transgene copy numbers in each cell. It is known that mature leaf cells in tobacco contain about 10,000 copies of chloroplast genomes per cell (Bendich, 1987). By virtue of achieving homoplasmy, it is inferred that there are 10,000 copies of transgenes per cell. Southern blots detected with the *mer*AB probe (2.3 kb in size) showed integration of specific genes, *mer*A and *mer*B, as a single fragment of 7.96 kb (Figs. 1A and 3D). The control untransformed tobacco plants and mutants did not show this fragment (Fig. 3D). If the *mer*AB probe would have detected any unexpected size fragments, it might be a nonspecific integration into other plant genomes (nuclear or mitochondria) as discussed elsewhere (Daniell and Parkinson, 2003), but this was not observed. The transgenic plants were fully characterized via PCR and Southern-blot analysis, which showed site-specific integration of the genes into the chloroplast genome and achievement of homoplasmy, even at very early stages of selection (T_0). No difference in homoplasmy was detected among plants transformed with the pLDR-MerAB or pLDR-MerAB-3′-UTR vector.

Total RNA from T_0 and T_1 plants transformed with the pLDR-MerAB-3′-UTR and the pLDR-MerAB was extracted and used to perform the northern-blot analysis with four different probes (the *mer*A, *mer*B, *mer*AB, and *aadA* probes). The *mer*A probe clearly showed the dicistron containing the *mer*B and *mer*A genes with sizes of 2,332 nucleotides and also a minor transcript for the *mer*A monocistron of 1,694 nucleotides (Fig. 4A). The *mer*B probe showed the *mer*AB dicistron (2,332 nucleotides) plus a less abundant transcript (1,448 nucleotides) containing the *aad*A and *mer*B genes, and the monocistron corresponding to the *mer*B (638 nucleotides) transcript (Fig. 4B). The *mer*AB probe helped to visualize different transcripts in a single blot, the *mer*B and *mer*A dicistronic transcript (2,332 nucleotides), the *mer*A monocistron (1,694 nucleotides), the *aad*A and *mer*B dicistron (1,448 nucleotides), and the *mer*B monocistron (638 nucleotides; Fig. 4C). The *aadA* probe showed transcripts for the dicistron containing the *aad*A and *mer*B genes and also the *aad*A monocistron of 810 nucleotides (Fig. 4D). The northern-blot analyses showed that the most abundant transcript is the dicistron (2,332 nucleotides) containing the *mer*A and *mer*B genes. Less abundant transcripts corresponding to the *aad*A/*mer*B dicistron (1,448 nucleotides), the *mer*A monocistron (1,694 nucleotides), the *mer*B monocistron (638 nucleotides), and to the *aad*A monocistron (810 nucleotides) were also detected. The high abundance of the *mer*AB dicistron in the pLDR-MerAB or the pLDR-MerAB-3′-UTR plants is an interesting observation. Contrary to the current dogma in the literature, these transcripts were stable even in the absence of a 3′-UTR believed to be required for transcript stability. In addition, there is an indication that processing occurs in between transgenes in transgenic chloroplasts even though no such processing sequences were engineered. Even though all three transgenes are transcribed from a single promoter, no tricistrons containing the *aad*A, *mer*B, and *mer*A genes were detected. Observed processing between transgenes might be due to recognition of bacterial intergenic sequences by the chloroplast protein synthesis machinery.

Bioassays

When 16-d-old tobacco plants were grown for 14 d in soil containing PMA concentrations of 0, 50, 100, and 200 μM, the merAB seedlings (pLDR-MerAB and pLDR-MerAB-3′-UTR clones) grew well at PMA concentrations up to 100 μM PMA, and survived the highest PMA concentration of 200 μM (Fig. 5). On the

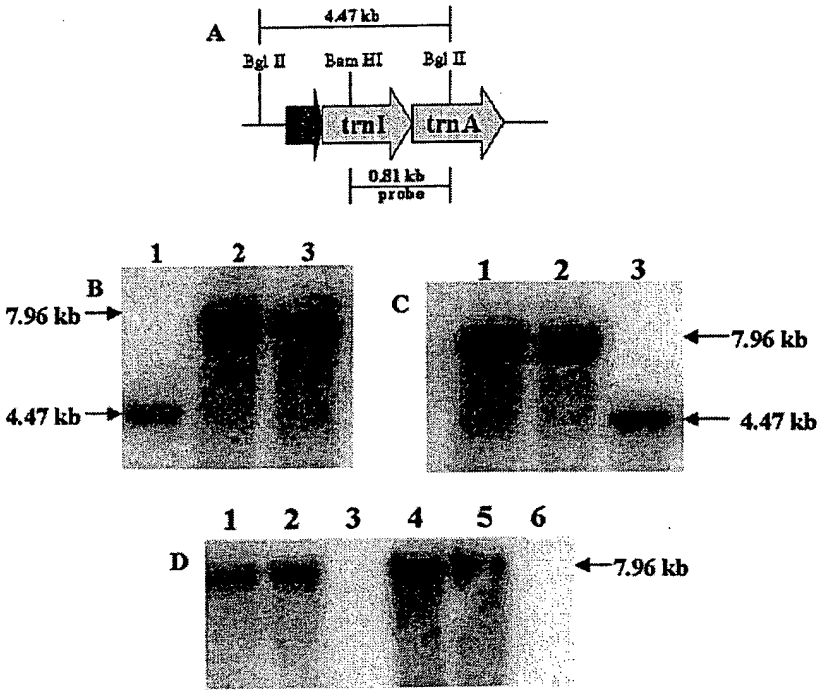


Figure 3. Southern-blot analysis using the flanking sequence probe and the *mer*AB probe. A, The map shows the wild-type chloroplast genome, restriction digestion sites used for Southern-blot analysis, and the 0.81-kb flanking sequence probe. B, Transgenic lines (T_0 generation) for the pLDR-MerAB (lane 2) and the pLDR-MerAB-3′-UTR (lane 3) show the expected size fragment of 7.96 kb; the untransformed control (lane 1) shows the 4.47-kb fragment. C, Lanes 1 and 2, T_1 generation transgenic lines; lane 3, the untransformed control. B and C, The flanking sequence probe was used. D, T_0 transgenic lines, pLDR-MerAB (lane 1), pLDR-MerAB-3′-UTR (lane 2), and their respective T_1 generation transgenic lines (lanes 4 and 5) show the 7.96-kb fragment. Lanes 3 and 6, Untransformed wild type. The *mer*AB probe was used in D.

other hand, PMA concentrations of 100 and 200 μM PMA were lethal to wild-type plants, which barely survived 50 μM PMA (Fig. 5). There were no significant differences between transgenic lines with or without the 3′-UTR terminator.

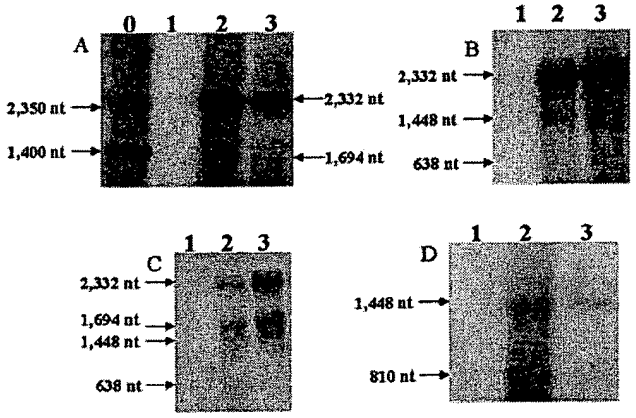


Figure 4. Northern-blot analysis. A, The *mer*A probe: transcripts of *mer*AB dicistron (2,332 nucleotides) and the *mer*A monocistron (1,694 nucleotides) are shown by arrows. B, The *mer*B probe: transcripts for the *mer*AB dicistron (2,332 nucleotides), the *aadA*/*mer*B dicistron (1,448 nucleotides), and the *mer*B monocistron (638 nucleotides) are shown. C, The *mer*AB probe: transcripts of the *mer*AB dicistron (2,332 nucleotides), the *mer*A monocistron (1,694 nucleotides), the *aadA*/*mer*B dicistron (1,448 nucleotides), and the *mer*B monocistron (638 bp) are shown. D, The *aadA* probe: transcripts of the *aadA*/*mer*B dicistron (1,448 nucleotides) and the *aadA* monocistron (810 nucleotides) are shown. 0, Marker; 1, wild-type, untransformed; 2, pLDR-MerAB transgenic line; 3, pLDR-MerAB-3′-UTR transgenic line.

The effect of PMA on plant growth was determined by treating 24-d-old tobacco plants with PMA concentrations of 0, 100, 200, 300, and 400 μM in soil and measuring total plant dry weight at each concentration (Fig. 6). The total dry weight of wild-type plants decreased progressively with each increase in PMA from 0 to 400 μM. On the other hand, in the transgenic plants, there was no decrease in total dry weight with increase in PMA concentration until PMA reached 400 μM. Statistical analysis (unpaired *t* test) showed that the transgenic lines were substantially more resistant than wild type to concentrations of PMA of 100, 200, and 400 μM (Table I). These results indicate clearly that, compared with the wild type, the insertion of *mer*A and *mer*B into the chloroplast genome substantially increased the resistance of the transgenic plants to the toxic effects of PMA. There was no significant difference between the dry weights of the two clones, pLDR-MerAB and pLDR-MerAB-3′-UTR, at each concentration of PMA tested (Fig. 6).

As discussed in the Introduction, previous research has shown that the main site of damage of organomercurial compounds is the chloroplast, and that chlorophyll synthesis, electron transport, and photosynthesis are all seriously affected. Therefore, the overexpression of *mer*A and *mer*B in the chloroplast should reduce the toxic effects of PMA directly on chloroplast function. To test this idea, we treated 15-mm diameter leaf discs excised from wild-type and transgenic plants with 10 μM PMA for 10 d and measured chlorophyll contents (Fig. 7). The results show that without PMA present, chlorophyll concen-

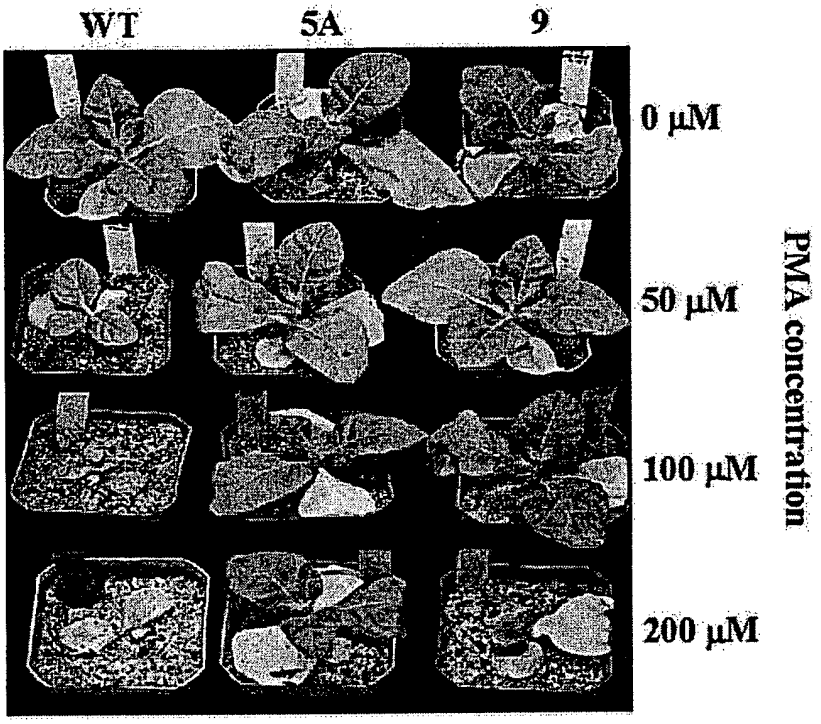


Figure 5. Effect of PMA concentration on the growth of wild-type and transgenic lines of tobacco plants. Seeds were germinated in vitro on Murashige and Skoog medium (without Suc and 0.5 g mL^{-1} spectinomycin). Seedling plants (10 d from germination) were transferred to a greenhouse and were grown in soil for 6 d. Plants were then treated by adding 200 mL of 0, 50, 100, and 200 μM PMA supplied in Hoagland nutrient solution. Photographs were taken 14 d after treatment. WT, Negative control cv Petit Havana; 5A, pLDR-MerAB transgenic line; 9, pLDR-MerAB-3′-UTR transgenic line.

tration did not differ between wild-type and the two transgenic lines. Surprisingly, when PMA was supplied to the leaf discs, the chlorophyll content was markedly increased in the transgenic lines, whereas in the wild type, chlorophyll content was reduced. These results are consistent with the view that PMA exerts a damaging effect on the chloroplasts of wild-type plants as expected, reducing chlorophyll content substantially, and that overexpression of *mer*A and *mer*B in the chloroplast genome appears to increase chloroplast resistance to PMA toxicity. However, because the overexpression of these genes results in an increase in chlorophyll content of the transgenic chloroplasts, it would appear that PMA could in fact stimulate chlorophyll synthesis in some way in these transgenic plants. In this regard, it is of interest that the leaf discs taken from the transgenic plants increased in size over the 10-d experimental period, whereas discs from the wild type decreased in size. Thus, it is possible that the increase in chlorophyll concentration with PMA in the transgenic plants was associated with an increase in chloroplast number and/or size.

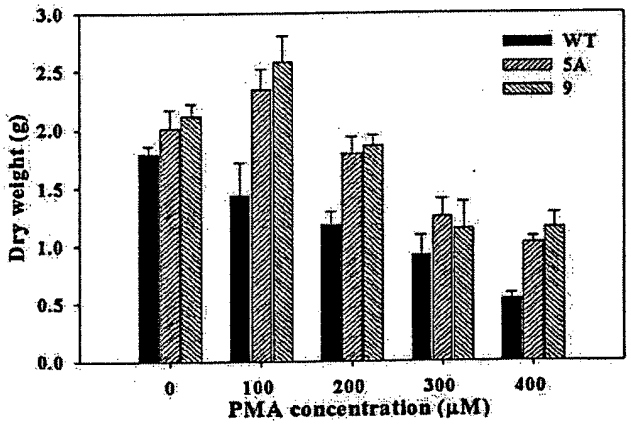


Figure 6. Effect of PMA on the total dry weight per plant of 24-d-old wild-type and transgenic tobacco plant lines grown on soil containing 0, 100, 200, 300, and 400 μM PMA for 14 d. WT, Negative control cv Petit Havana; 5A, pLDR-MerAB transgenic line; 9, pLDR-MerAB-3′-UTR transgenic line. SE shown, $n = 5$.

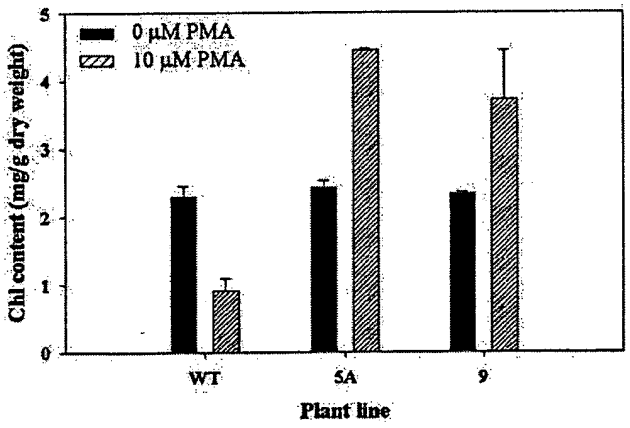


Figure 7. Effect of PMA on total chlorophyll content (milligrams per gram of dry weight) of 15-mm diameter leaf discs excised from wild-type and transgenic lines of tobacco and treated with 0 and 10 μM PMA for 6 d. WT, Negative control cv Petit Havana; 5A, pLDR-MerAB transgenic line; 9, pLDR-MerAB-3′-UTR transgenic line. SE shown, $n = 5$.

Table I. *Unpaired t test values comparing the differences in dry weight between each transgenic line of tobacco versus wild type*

5A: pLDR-MerAB transgenic line; 9: pLDR-MerAB-3'UTR transgenic line. An asterisk indicates significance at $P < 0.05$; a double asterisk indicates significance at $P < 0.001$.

	PMA Concentration									
	Control		100		200		300		400	
	5A	9	5A	9	5A	9	5A	9	5A	9
					μM					
Dry weight	1.31	2.60*	2.77*	3.19**	3.38**	4.62**	1.41	0.79	6.67**	4.72**

Levels of transgene expression in chloroplasts could be further enhanced by introducing appropriate UTRs instead of the ribosome-binding site (RBS) used in the present study. For example, we have recently shown that use of the *psb*A 5'-UTR instead of RBS resulted in a 500-fold increase in the expression of human serum albumin in transgenic chloroplasts (Fernandez-San Millan et al., 2003). The most significant advantage is the ability to introduce the *mer* operon in a single transformation event in contrast to nuclear transgenic plants that required introduction of single genes followed by time consuming backcrosses to reconstitute the entire pathway. In addition, prokaryotic genes do not require codon optimization when expressed in transgenic chloroplasts (Kota et al., 1999; De Cosa et al., 2001).

This is the first report on the use of chloroplast transformation using multigene engineering for the phytoremediation of toxic compounds. Because of the containment of transgenes and high levels of expression via chloroplast genomes, the chloroplast transformation approach is highly suitable for phytoremediation, especially for toxic agents that affect chloroplast function. Although 3'-UTR is believed to stabilize chloroplast transcripts and to be essential for transgene expression, it may not be necessary for transcript stability in the context of a polycistron. Because there are more than 60 such polycistrons within the chloroplast genome (Sugita and Sugiura, 1996), this is a significant observation.

MATERIALS AND METHODS

Bacterial Plasmids That Contain Organomercurial and Hg Resistance Genes

Host *Escherichia coli* cells containing plasmids NR1 and R831b were kindly provided by Dr. Ann Summers (University of Georgia, Athens). These plasmids contain the mer operon with the complete and functional merA and merB genes, respectively (Jackson and Summers, 1982; Rinderle et al., 1983; Ogawa et al., 1984; Begley et al., 1986). Each of these plasmids confers resistance to at least one antibiotic that can be used as a selectable marker. Host bacterial containing plasmid NR1 was grown on solid Luria-Bertani media containing 100 $\mu g\ mL^{-1}$ tetracycline; *E. coli* cells containing the plasmid R831b was cultured on solid Luria-Bertani media containing 12.5 $\mu g\ mL^{-1}$ kanamycin and were grown overnight at 37°C.

Chloroplast Vector Constructions

To amplify the *mer*B gene from the native plasmid, a primer pair was designed to have a *Pst*I restriction site followed by a chloroplast and bacterial functional RBS of sequence GGAGG in the 5' primer, followed by a four-nucleotide spacer region upstream of the start codon. This primer had 20-nucleotide homology with the 5' end of the gene and a total of 35 nucleotides. The 3' primer was designed to have 20-nucleotide homology with the 3' end of the gene and a *Cla*I restriction site. To amplify the *mer*A gene from the native plasmid, a 5' primer was designed to have a *Cla*I restriction site followed by the RBS sequence and a four-nucleotide spacer region before the start codon and the 20-nucleotide homology with the *mer*A gene. All primer pairs were designed using the QUICKPRI program of the DNASTAR software. Two PCR reactions were done to amplify the *mer*A and the *mer*B genes individually from the plasmid NR1 that contained the complete and functional *mer*A gene and the plasmid R831b that contained the full-length *mer*B gene. The PCR products were cloned into suitable plasmid vectors.

pLDR-MerAB-3'-UTR Vector Construction

The functional *mer*AB operon was amplified via PCR from the vector pCR2.1-MerAB and a new set of primers was made. The 5' primer was designed to have an *Eco*RV site, an RBS, a spacer region of four nucleotides (attt) and 20 bases of homology to the *mer*AB operon starting at the start codon (atg). The 3' primer is a simple primer with 20 bases of homology to the 3' end of the operon. After cloning, correct orientation was verified by restriction analyses.

Hg Resistance Bioassay in Bacteria

The bacterial clones pLDR-MerAB, pLDR-MerAB-3'-UTR, and the control *E. coli* XL1-blue cells were grown for 24 h at 37°C in 50 mL of Luria-Bertani broth with concentrations of $HgCl_2$ of 0, 25, and 50 μM. The growth medium was autoclaved and cooled to 40°C before adding $HgCl_2$, and was mixed thoroughly to provide an even concentration throughout the plate or growth medium. The bacterial clones pLDR-MerAB, pLDR-MerAB-3'-UTR, and the untransformed control *E. coli* cells were plated in solid Luria-Bertani medium containing $HgCl_2$ concentrations of 0, 50, 100, and 500 μM. Plates were incubated for 24 h at 37°C.

Bombardment and Selection of Transgenic Plants

The steps involved in the gene delivery by particle bombardment and the selection process of the transgenic tobacco (*Nicotiana tabacum* var Petit Havana) clones were performed essentially as describe by Daniell (1997). Tobacco leaves were bombarded using a biolostic device (PDS-1000/He; Bio-Rad, Hercules, CA). After bombardment, leaves were placed on Regeneration Medium of Plants medium containing 500 $\mu g\ mL^{-1}$ spectinomycin for two rounds of selection on plates and subsequently moved to jars on Murashige Skoog medium containing 500 $\mu g\ mL^{-1}$ spectinomycin.

Confirmation of Chloroplast Integration by PCR

Plant DNA was isolated using the DNeasy Plant Mini kit (Qiagen, Valencia, CA). The PCR primer pairs 3P-3M and 5P-2M were used to confirm the integration of the gene cassette into the chloroplast and the presence of the genes of interest, respectively, essentially as described elsewhere (Guda et al., 2000). PCR analysis was performed using the Gene Amp PCR System 2400 (Perkin Elmer, Chicago).

Southern-Blot Analysis

The total plant DNA was obtained from transgenic T_0 and T_1 plants as well as from untransformed tobacco plants following the protocol previously explained (Daniell et al., 2001a,b). The plant DNA was digested with *Bgl*II and was separated on a 0.8% (w/v) agarose gel at 50 V for 2 h. The gel was soaked in 0.25 N HCl for 15 min and was then rinsed two times with water. The gel was then soaked in transfer buffer (0.4 N NaOH and 1 M NaCl) for 20 min and transferred overnight to a nitrocellulose membrane. The membrane was rinsed twice in 2× SSC (0.3 M NaCl and 0.03 M sodium citrate), dried on filter paper, and then cross-linked in the GS GeneLinker (Bio-Rad). The flanking sequence probe was obtained by *Bgl*II/*Bam*HI digestion of the plasmid pUC-ct that contains the chloroplast-flanking sequences (*trn*I and *trn*A genes). The *mer*AB probe was obtained by *Eco*RI digestion of plasmid pCR2.1-MerAB. Probes were labeled with ^{32}P using Ready Mix and were purified by using Quant G-50 microcolumns (Amersham, Arlington Heights, IL), followed by radioisotope incorporation. The probe was quantified by using a scintillation counter (LS 5000TD; Beckman Instruments, Fullerton, CA). Prehybridization and hybridization were done using the Quick-Hyb solution (Stratagene, La Jolla, CA). The membrane was washed twice in 2× SSC with 0.1% (w/v) SDS for 15 min at room temperature, followed by two additional washes in 0.1× SSC with 0.1% (w/v) SDS for 15 min at 60°C (to increase the stringency). Blots were exposed to x-ray films and were developed in a SRX-101A (Konica, Tokyo).

Northern-Blot Analysis

The RNeasy Mini kit and protocol was used to isolate total RNA from plant tissues (Qiagen). The *mer*A, *mer*B, *aad*A, and *mer*AB probes were used to probe different RNA blots. The *mer*A probe was made by cutting out the *mer*A gene from the pCR2.1-MerA vector with *Eco*RI. The *mer*B probe was made by cutting out the *mer*B gene from the pCR2.1-MerB vector with *Eco*RI. The *aad*A probe was amplified by PCR from the pLD-ctv vector with a specific primer pair (5′-ccatggcagaagcggtaatcg/3′-aagatttatttgccgactacctt). The *mer*AB probe was made digesting the pCR2.1-MerAB vector with *Eco*RI. Restriction fragments were cut out and eluted from the gels. The probe-labeling reaction, prehybridization/hybridization steps, membrane washing step, and autoradiography were performed as explained in the Southern-blot section in "Materials and Methods."

PMA Treatments

Seeds of wild-type tobacco and two transgenic lines (pLDR-MerAB and pLDR-MerAB-3′-UTR) were surface-sterilized in 7% (w/v) sodium hypochlorite containing 0.1% (v/v) Tween 20. Seeds were kept on a rocking platform for 20 min and were rinsed in sterile distilled water at least three times. Sterilized seeds were transferred to plates containing one-half-strength Murashige and Skoog medium (Murashige and Skoog, 1962) with 0.5 mg mL^{-1} spectinomycin and 0.3% (w/v) phytoagar, pH 5.7. Plates were incubated in the dark at 4°C for 3 d, and were then maintained in a controlled growth chamber at a temperature of 22°C to 24°C, relative humidity of 75% to 90%, and a photon flux density of 750 $\mu E\ m^{-2}$ supplied over a 16-h daylength. After germination (approximately 10 d), seedlings were transferred to soil (sand:Davis Mix, 50:50) in the greenhouse at 22°C using a 16-h photoperiod. Five replicate pots each contained a single seedling, wild-type or transgenic plant. All pots were watered twice a week with one-half-strength Hoagland solution.

Effect of PMA on Seedling Germination

To determine the inhibitory concentration of PMA on seedling germination, three different concentrations of PMA were applied to pots containing 16-d-old plants from wild-type and two transgenic lines in three replicates. PMA stock solutions were prepared as 10 mM in dimethyl sulfoxide. Different PMA concentrations (50–200 μM) were added to each pot in 100 mL of one-half-strength Hoagland solution. Control pots received the same volume of Hoagland solution without PMA. All plants were grown in the greenhouse under the same conditions as described above.

Effect of PMA on Potted Plants

Pots of five replicates representing the wild-type and the two transgenic lines (of approximately the same size) were transferred to Poly Vinyl Chloride plastic trays 3 inches high. Different concentrations of PMA (in micromoles) were prepared (100, 200, 300, and 400) using a stock solution of one-half-strength Hoagland solution. For each treatment, a single tray maintained approximately 200 mL (to about one-half of the pot's height) of the PMA-Hoagland's solution. All plants in the same treatment were exposed to exactly the same concentration of PMA. The control tray was filled with one-half-strength Hoagland solution without metal. After about 14 d, plants were harvested, washed thoroughly with distilled water, and the length of the longest root and shoot of the plants were measured. Shoots and roots were separated and dry weights were determined.

Determination of Chlorophyll Content in Leaf Discs Treated with PMA

Leaf discs were cut out with a cork-borer (15-mm diameter) from the youngest and fully expanded leaves on 3-week-old plants grown in the soil with no PMA. Discs of wild-type and different transgenic plants were placed in petri dishes containing solidified Murashige and Skoog medium (pH 5.7 with no Suc) supplemented with different concentrations of PMA ranging from 0.1 to 1 μM, 10 to 100 μM, and 200 to 500 μM. Plates with no PMA were used as controls. The effect of Hg stress was assessed by the loss of chlorophyll in leaf discs. Leaf discs were collected after 6 d of exposure to PMA. They were immediately extracted in 80% (v/v) chilled acetone for determination of total chlorophyll content following the protocol from Current Protocols in Food Analytical Chemistry Online (http://www.mrw2.interscience.wiley.com).

ACKNOWLEDGMENT

We thank Dr. Ann Summers (University of Georgia) for providing bacterial strains used in this study.

Received January 24, 2003; returned for revision March 4, 2003; accepted April 1, 2003.

LITERATURE CITED

Begley TP, Walts AE, Walsh CT (1986) Bacterial organomercurial lyase: overproduction, isolation, and characterization. Biochemistry 25: 7186–7192

Bendich AJ (1987) Why do chloroplasts and mitochondria contain so many copies of their genome? Bioassays 6: 279–282

Bernier M, Popovic R, Carpentier R (1993) Mercury inhibition of photosystem II. FEBS Lett 32: 19–23

Bizily S, Rugh CC, Meagher RB (2000) Phytoremediation of hazardous organomercurials by genetically engineered plants. Nat Biotechnol 18: 213–217

Bizily S, Rugh CC, Summers AO, Meagher RB (1999) Phytoremediation of methylmercury pollution: merB expression in *Arabidopsis thaliana* plants confer resistance to organomercurial Proc Natl Acad Sci USA 96: 6808–6813

Daniell H (1997) Transformation and foreign gene expression in plants mediated by microprojectile bombardment. Methods Mol Biol 62: 453–488

Daniell H (2002) Molecular strategies for gene containment in transgenic crops. Nat Biotechnol 20: 581–586

Daniell H (2003) Medical molecular pharming: expression of antibodies biopharmaceuticals and edible vaccines via the chloroplast genome. In K Vasil, ed, Plant Biotechnology 2002 and Beyond. Kluwer Academic Publisher, Dordrecht, The Netherlands, pp 371–376

Daniell H, Datta R, Varma S, Gray S, Lee SB (1998) Containment of herbicide resistance through genetic engineering of the chloroplast genome. Nat Biotechnol 16: 345–348

Daniell H, Dhingra A (2002) Multigene engineering: dawn of an exciting new era in biotechnology. Curr Opin Biotechnol 13: 136–171

Daniell H, Khan MS, Allison L (2002) Milestones in chloroplast genetic engineering: an environmental friendly era in biotechnology. Trends Plant Sci 7: 84–91

Daniell H, Lee SB, Panchal T, Weibe PO (2001a) Expression of the native cholera toxin B subunit gene and assembly as functional oligomers in transgenic chloroplasts. J Mol Biol **311:** 1001–1009

Daniell H, Muthukumar B, Lee SB (2001b) Marker free transgenic plants: engineering the chloroplast genome without the use of antibiotic selection. Curr Genet **39:** 109–116

Daniell H, Parkinson L (2003) Jumping genes and containment. Nat Biotechnol **21:** 374–375

Dean JG, Bosqui FL, Lanouette VH (1972) Removing heavy metals from waste water. Environ Sci Technol **6:** 518–522

De Cosa B, Moar W, Lee SB, Miller M, Daniell H (2001) Hyper-expression of the Bt Cry2Aa2 operon in chloroplasts leads to formation of insecticidal crystals. Nat Biotechnol **19:** 71–74

De Gray G, Rajasekaran K, Smith F, Sanford J, Daniell H (2001) Expression of an antimicrobial peptide via chloroplast genome control phytopathogenic bacteria and fungi. Plant Physiol **127:** 852–862

Doucleff M, Terry N (2002) Pumping out the arsenic. Nat Biotechnol **20:** 1094–1095

Fernandez-San Millan A, Mingo-Castel A, Miller M, Daniell H (2003) A chloroplast transgenic approach to hyper-express and purify human serum albumin, a protein highly susceptible to proteolytic degradation. Plant Biotechnol J **1:** 71–79

Guda C, Lee SB, Daniell H (2000) Stable expression of biodegradable protein base polymer in tobacco chloroplasts. Plant Cell Rep **19:** 257–262

Heaton ACP, Rugh CL, Wang NJ, Meagher RB (1998) Phytoremediation of mercury and methylmercury polluted soils using genetically engineered plants. J Soil Contam **7:** 497–509

Jackson WJ, Summers AO (1982) Biochemical characterization of $HgCl_2$-inducible polypeptides encoded by the mer operon of plasmid R100. J Bacteriol **151:** 962–970

Kärenlampi S, Schat H, Vangronsveld J, Verkleij JAC, Lelie D, Mergeay M, Tervahauta AI (2000) Genetic engineering in the improvement of plants for phytoremediation of metal polluted soils. Environ Pollut **107:** 225–231

Kota M, Daniell H, Varma S, Garczynski F, Gould F, Moar WJ (1999) Overexpression of the *Bacillus thuringiensis* Cry2A protein in chloroplast confers resistance to plants against susceptible and Bt-resistant insects. Proc Natl Acad Sci USA **96:** 1840–1845

Krämer U, Chardonnens AN (2001) The use of transgenic plants in the bioremediation of soil contaminated with trace elements. Appl Microbiol Biotechnol **55:** 661–672

Kupper H, Kupper F, Spiller M (1996) Environmental relevance of heavy metal substituted chlorophylls using the example of water plants. J Exp Bot **47:** 259–266

Lee SB, Kwon S, Park S, Jeong M, Han S, Byun M, Daniell H (2003) Accumulation of trehalose within transgenic chloroplasts confers drought tolerance. Mol Breed **11:** 1–13

Lin ZQ, Hansen D, Zayed AM, Terry N (1995) Biological selenium volatilization: method of measurement under field conditions. J Environ Qualt **28(1):** 309–315

McBride KE, Svab Z, Schaaf DJ, Hogen PS, Stalker DM, Maliga P (1995) Amplification of a chimeric Bacillus genes in chloroplasts leads to an extraordinary level of an insecticidal protein in tobacco. BioTechnology **13:** 362–365

Meagher RB (2000) Phytoremediation of toxic elemental and organic pollutants. Curr Opin Plant Biol **3:** 153–162

Meagher RB, Rugh CL (1997) Phytoremediation of heavy metal pollution: ionic and methyl mercury. In Organization for Economic Cooperation and Development Document: Biotechnology for Water Use and Conservation. The Mexico '96 Workshop, Organization for Economic Cooperation and Development, Paris, pp 305–321

Murashige T, Skoog F (1962). A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Planta **15:** 473–497

Nies DH (1999) Microbial heavy metal resistance. Appl Microbiol Biotechnol **51:** 730–750

Ogawa HI, Tolle CL, Summers AO (1984) Physical and genetic map of the organomercury resistance (Omr) and inorganic mercury resistance (Hgr) loci of the IncM plasmid R831b. Gene **32:** 311–320

Patra M, Sharma A (2000) Mercury toxicity in plants. Bot Rev **66:** 379–422

Prasad DDK, Prasad ARK (1987) Altered *o*-aminolevulinic acid metabolism by lead and mercury in germination seedlings of bajra (*Pennisetum typoideum*). J Plant Physiol **127:** 241–249

Rinderle SJ, Booth JE, Williams JW (1983) Mercuric reductase from R-plasmid NR1: characterization and mechanistic study. Biochemistry **22:** 869–876

Rugh CL, Senecoff JF, Meagher RB, Merkle SA (1998) Development of transgenic yellow poplar for mercury phytoremediation. Nat Biotechnol **16:** 925–928

Rugh CL, Wilde HD, Stack NM, Thompson DM, Summers AO, Meagher RB (1996) Mercuric ion reduction and resistance in transgenic *Arabidopsis thaliana* plants expressing a modified bacterial *merA* gene. Proc Natl Acad Sci USA **93:** 3182–3187

Salt DE, Smith RD, Raskin I (1995) Phytoremediation: a novel strategy for the removal of toxic metals from the environment using plants. BioTechnology **13:** 468–474

Sen AK, Mondal NG (1987) *Salvia natans* as the scavenger of Hg (II). Water Air Soil Pollut **34:** 439–446

Sinha S, Gupta M, Chandra P (1996) Bioaccumulation and biochemical effects of mercury in the plant *Bacopa monnieri* (L). Environ. Toxicol Water Qualt **11:** 105–112

Staub JM, Garcia B, Graves J, Hajdukiewicz PTJ, Hunter P, Nehra N, Paradkar V, Schlittler M, Carol JA, Spatola L et al. (2000) High-yield production of human therapeutic protein in tobacco chloroplast. Nat Biotechnol **18:** 333–338

Sugita M, Sugiura M (1996) Regulation of genes expression in chloroplast of higher plants. Plant Mol Biol **32:** 315–326

Terry N, Zayed AM, De Souza MP, Tarun AS (2000) Selenium in higher plants. Annu Rev Plant Physiol Plant Mol Biol **51:** 401–432

U.S. Environmental Protection Agency (1984) Health effects assessment of mercury. Environmental Criteria and Assessment Office, Cincinnati, OH

RUF *ET AL.* 2001

RESEARCH ARTICLE

Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit

Stephanie Ruf[1,2], Marita Hermann[1], Irving J. Berger[3], Helaine Carrer[3], and Ralph Bock[1,2*]

Transgenic chloroplasts offer unique advantages in plant biotechnology, including high-level foreign protein expression, absence of epigenetic effects, and gene containment due to the lack of transgene transmission through pollen. However, broad application of plastid genome engineering in biotechnology has been largely hampered by both the lack of chloroplast transformation systems for major crop plants and the usually low plastid gene expression levels in nongreen tissues such as fruits, tubers, and other storage organs. Here we describe the development of a plastid transformation system for tomato, *Lycopersicon esculentum*. This is the first report on the generation of fertile transplastomic plants in a food crop with an edible fruit. We show that chromoplasts in the tomato fruit express the transgene to ~50% of the expression levels in leaf chloroplasts. Given the generally very high foreign protein accumulation rates that can be achieved in transgenic chloroplasts (>40% of the total soluble protein), this system paves the way to efficient production of edible vaccines, pharmaceuticals, and antibodies in tomato.

The genetic information of plants is distributed among three cellular compartments: the nucleus, the mitochondria, and the plastids. Each of these compartments carries its own genome and, consequently, expresses heritable traits. Plastids and mitochondria do not obey the mendelian rules and usually exhibit uniparental transmission to the next generation. The plastid genome of higher plants is a circular double-stranded molecule of 120 to 160 kilobases, harboring ~130 genes[1]. Identical copies of this genome are present in all cells and all plastid types (e.g., undifferentiated proplastids, photosynthesis-performing chloroplasts, carotenoid-accumulating chromoplasts, and starch-storing amyloplasts). A remarkable feature of the plastid genome is its extremely high ploidy level: a single tobacco leaf cell may contain as many as 100 chloroplasts, each harboring ~100 identical copies of the plastid genome, resulting in an extraordinarily high ploidy degree of up to 10,000 plastid genomes per cell[2].

The development of technologies to engineer the chloroplast genome of the green alga *Chlamydomonas reinhardtii*[3] and the higher plant *Nicotiana tabacum*[4] has opened up the possibility to target transgenes to the plastid genome by chloroplast transformation. These technologies offer a great potential for the biotechnology of the future[5–7] and a number of most attractive advantages over conventional transgenic plants (generated by transformation of the nuclear genome), such as (1) high levels of transgene expression and foreign protein accumulation of up to >40% of the total soluble cellular protein (presumably resulting from the polyploidy of the plastid genetic system and/or the high stability of foreign proteins)[8–10]; (2) the possibility of expressing multiple transgenes as operons ("transgene stacking") due to efficient translation of polycistronic messenger RNAs (mRNAs) in plastids[11]; (3) absence of position effects in plastids due to lack of a compact chromatin structure and efficient transgene integration by homologous recombination[12]; (4) absence of epigenetic effects (gene silencing); and (5) transgene containment due to uniparentally maternal inheritance of chloroplasts in most higher plants (i.e., absence of pollen transmission of transgenes)[13–15].

In higher plants, chloroplast transformation is routinely available only in tobacco, *N. tabacum*. The main obstacle to extending the technology to other species and, most importantly, to major crops is probably posed by limitations in the currently available tissue culture systems and regeneration protocols for transplastomic plants. Although some progress was made recently with *Arabidopsis* and potato chloroplast transformation[16,17], the production of fertile transplastomic plants in any other species but tobacco has not yet been reported (reviewed in ref. 18). In fact, the three chloroplast transformants generated to date for the model plant *Arabidopsis thaliana* all were sterile and hence could not be propagated generatively[16].

We report here the development of a stable plastid transformation system for tomato. Transplastomic tomato plants produced fruits and viable seeds, which transmitted the transgene in a uniparentally maternal fashion as expected for a plastid-encoded trait. Moreover, we observe efficient transgene expression not only in chloroplasts of green leaves but also in chromoplasts of tomato fruits, demonstrating the potential of the system for biotechnological applications.

Results

Construction of a new versatile vector series for plastid transformation. In chloroplast transformation experiments with tobacco, we tested whether numerous regions of the chloroplast genome were suitable target sites for the uptake of transgenes. In the course of this work, we identified a region in the chloroplast genome (Fig. 1A) that, when used in transformation vectors as targeting sequence for homologous recombination, resulted in particularly high chloroplast transformation frequencies (in three independent experiments with 30 bombardments each, plastid transformation frequencies were between 1.5 and 4 times higher than with control vectors of the

[1]*Institut für Biologie III, Universität Freiburg, Schänzlestraße 1, D-79104 Freiburg, Germany.* [2]*Institut für Biochemie und Biotechnologie der Pflanzen, Westfälische Wilhelms-Universität Münster, Hindenburgplatz 55, D-48143 Münster, Germany.* [3]*Centro de Biotecnologia Agrícola, ESALQ, Universidade de São Paulo, Av. Pádua Dias 11, Piracicaba, SP, 13418-900, Brazil.* *Corresponding author (rbock@uni-muenster.de).

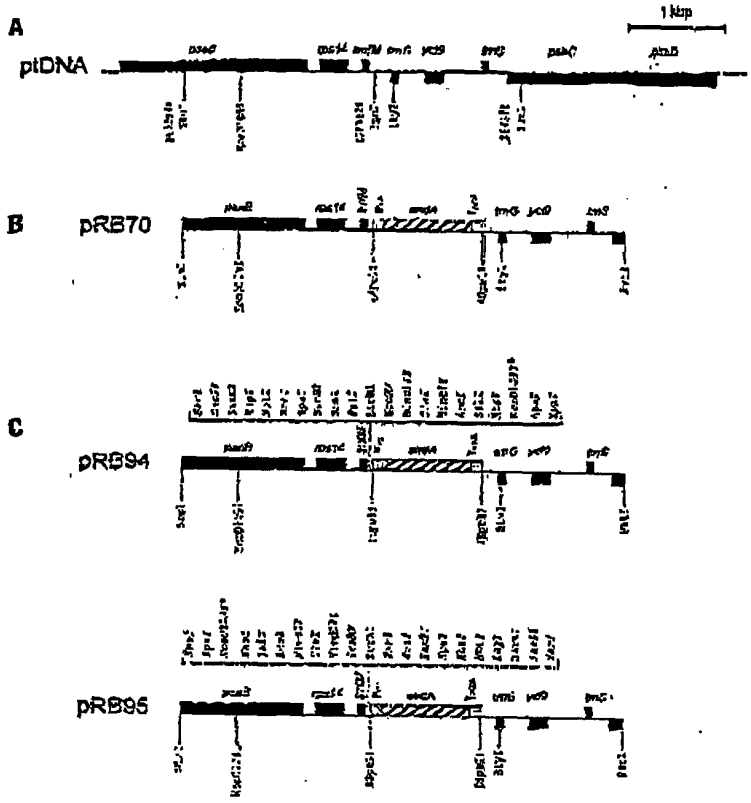


Figure 1. Construction of chloroplast transformation vectors containing polylinkers for cloning of passenger genes. (A) Physical and restriction map of the chloroplast genome region used for construction of plasmid vectors for plastid transformation. Genomic nucleotide positions defining the plastid targeting fragment used for vector construction (*Stu*I and *Pst*I sites) as well as the *aadA* insertion site (*Spe*I site) are given (according to ref. 34). (B) Map of the plastid transformation vector pRB70 (EMBL accession no. AJ312391) containing a chimeric *aadA* gene driven by the *rRNA* operon promoter. (C) Map of vectors pRB94 (EMBL accession no. AJ312392) and pRB95 (EMBL accession no. AJ312393). The two plasmids carry the polylinker taken from pBluescript upstream of the *aadA* gene in different orientations. Several restriction sites for polylinker enzymes within the vector sequences were eliminated by mutagenesis. As a result, almost all restriction sites present in the polylinker are unique (enzymes shown in italics) and thus can be used for insertion of passenger genes. Note that the *Eco*O109I site in the polylinker, although not unique, can be utilized for cloning in that the second site within the *psaB* gene is Dcm methylated in *Escherichia coli* and hence not recognized in plasmids prepared from standard (*dcm*$^+$) laboratory strains. Restriction sites eliminated by mutagenesis or ligation of heterologous ends are shown in parentheses.

pRB51 series[19]). We chose this region and derived from it a series of plastid transformation vectors. In these constructs, the selectable marker gene *aadA* was inserted between two transfer RNA (tRNA) genes (Fig. 1B,C). For the convenient insertion of passenger genes and their tight linkage to the selectable spectinomycin resistance gene *aadA*, we inserted the pBluescript polylinker in different orientations immediately upstream of the *aadA* marker (Fig. 1C).

All constructs were first tested in tobacco, where they produced chloroplast transformants at high frequency (data not shown). The plastid targeting region in the transformation vectors is highly conserved in the chloroplast genomes of dicotyledonous plants and hence is expected to be suitable not only for plastid transformation in the two closely related solanaceous species tobacco and tomato but also for other higher plants. An exception are graminaceous species, such as rice and corn, which (1) carry a large inversion with one of the breakpoints being in the *orf44/rps14* region[20,21] and (2) have a different RNA editing pattern of the *rps14* transcript (only one out of the two editing sites present in tobacco is conserved in maize)[21,22].

Generation of tomato plants with transgenic plastids. The keys to the successful transformation of tomato chloroplasts were (1) the use of extreme low-light conditions during the entire selection phase; (2) the drastic extension of the primary selection phase to three or four months compared with three to five weeks in tobacco; (3) the significantly smaller size of the leaf pieces exposed to the tissue culture medium during primary selection as compared to tobacco plastid transformation (optimal size 3 × 3 mm; Fig. 2A); and (4) the optimization of the selection and plant regeneration scheme (see Experimental Protocol and below).

Plastid transformation experiments were carried out by biolistic bombardment of sterile tomato leaves with plasmid pRB70. Primary spectinomycin-resistant calli were selected after three to four months' incubation of bombarded leaf pieces on a plant tissue culture medium containing spectinomycin (300 or 500 μg/ml). Putative transplastomic calli were yellow or pale green (Fig. 2A) and appeared to be very sensitive to light at this stage. In general, successful selection of transplastomic tomato cells was found to be critically dependent on the use of much lower light intensities (25 μE) than for selection of transplastomic tobacco plants (70–100 μE). When we used our standard conditions for tobacco plastid transformation experiments (5 × 5 mm size of the leaf pieces subjected to selection; ~90 μE light intensity), we recovered neither transplastomic tomato calli nor spontaneous spectinomycin-resistant lines.

For further propagation and generation of homoplasmic cells that lack any residual copies of the wild-type plastid genome, tissue samples from primary spectinomycin-resistant calli were transferred onto the surface of fresh culture medium with spectinomycin[4]. Unlike tobacco, tomato tissue does not show shoot development on this medium and keeps growing as green calli (Fig. 2B). At this stage, successful chloroplast transformation was verified. An initial fast test by PCR identified chloroplast transformants and

Table 1. Results of three independent chloroplast transformation experiments in tomato

Experiment no.	Bombarded leaf samples	Number of selection plates	Spectinomycin-resistant calli[a]	PCR Positive	RFLP Positive
1	20	180	6	3	3
2	20	180	3	1	1
3	20	180	4	2	2

[a]Calli displaying resistance to spectinomycin but being negative in PCR and RFLP tests are likely to be spontaneous resistance mutants that arise through acquisition of point mutations in the 16S rRNA gene[36].

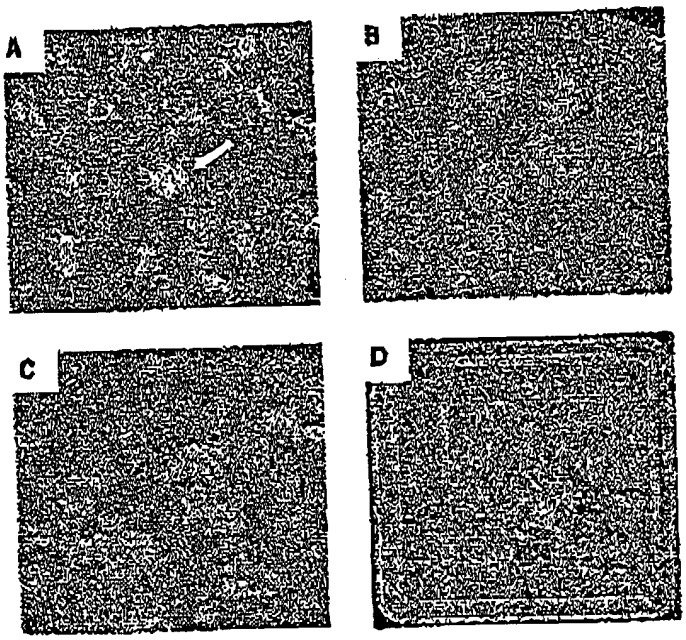


Figure 2. Generation of tomato plants with transgenic plastids. (A) Primary selection of spectinomycin-resistant tomato calli. A plate with bombarded leaf pieces is shown after three months' incubation on spectinomycin-containing tissue culture medium. Note that the leaf tissue is bleached out as a result of effective inhibition of chloroplast translation. Spectinomycin-resistant calli appear as small yellow or pale green mounts of dividing cells (arrow). (B) Propagation of spectinomycin-resistant tomato lines. Tissue samples from primary plastid transformants are subjected to additional selection cycles on antibiotic-containing culture medium. On this medium, the tissue grows as undifferentiated callus from which samples are taken in regular intervals for homoplasmy tests (Fig. 3). (C) Plant regeneration from homoplasmic transplastomic callus tissue. Shoot regeneration is observed approximately four weeks after transfer of homoplasmic callus material to shoot induction medium. (D) Rooting of transplastomic tomato shoots. Shoots induced from homoplasmic calli are transferred to boxes with hormone-free rooting medium. Following successful rooting, plants are transferred to the soil and grown to maturity in the greenhouse.

allowed elimination of spontaneous spectinomycin-resistant lines (Table 1). After one to two additional cycles of callus propagation on the identical medium, chloroplast transformation and homoplasmy were ultimately confirmed by restriction fragment length polymorphism (RFLP) analyses (Fig. 3). As the *aadA* gene confers broad-range resistance to a variety of antibiotics of the aminoglycoside type[12], we additionally tested chloroplast transformants for double resistance to both spectinomycin and streptomycin. Whereas spontaneous spectinomycin-resistant lines are sensitive to streptomycin and bleach out on tissue culture medium containing these two drugs, continued callus growth of transplastomic tissue (data not shown) provided further evidence of successful chloroplast transformation and efficient expression of the plastid *aadA* marker.

For plant regeneration, homoplasmic callus tissue was placed onto the surface of shoot induction medium containing indole-3-acetic acid (IAA) as auxin and either 6-benzylaminopurine (BAP) or zeatin as cytokinin (Fig. 2C). For subsequent rooting, shoots were transferred into boxes containing phytohormone-free medium (Fig. 2D). The resulting transplastomic plants were then planted into the soil and grown to maturity in the greenhouse.

To confirm uniparentally maternal transgene transmission to the next generation, we pollinated emasculated flowers from transplastomic plants with pollen from wild type plants. As expected for a plastid-encoded trait, the F1 progeny resulting from these crosses was uniformly spectinomycin-resistant (Fig. 4) confirming both stable transformation of the tomato plastid genome and homoplasmy of the transplastomic lines. In addition, molecular analysis of F1 progeny plants confirmed presence of the *aadA* transgene in a homoplasmic state (see Fig. 6A).

A single bombarded tomato leaf sample typically resulted in eight to nine selection plates with leaf pieces on the surface of spectinomycin-containing callus induction medium (Table 1). In three independent transformation experiments, we selected altogether six tomato chloroplast transformants, equaling transformation efficiency of one transplastomic line selected from approximately 80–100 selection plates. Although this efficiency is significantly lower than the plastid transformation frequency in the well-established tobacco system (where one bombarded leaf sample typically results in four selection plates and we routinely obtain one chloroplast transformant per 5–10 selection plates), tomato plastid transformation is efficient enough to provide a workable system for both basic research and plant biotechnology.

Plastid transgene expression in tomato leaves and fruits. The promoter used to drive plastid transgene expression in tomato is the strong ribosomal RNA (rRNA) operon promoter P_{rrn} (ref. 13). In tobacco, this promoter was shown to confer foreign protein accumulation of up to 5% of the total soluble leaf protein[3] and in one case

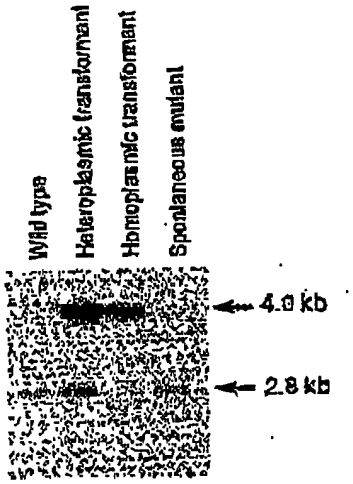


Figure 3. RFLP analysis to identify homoplasmic transplastomic tomato lines. A wild-type sample, a spontaneous spectinomycin-resistant line, a heteroplasmic transplastomic line, and a homoplasmic line are shown. DNAs were digested with *Eco*O109I and *Pst*I and probed with a radiolabeled *Sty*I/*Pst*I restriction fragment (Fig. 1). To increase the sensitivity of the assay, approximately five times more DNA from the transplastomic lines was loaded. Whereas the heteroplasmic line clearly contains a mix of wild-type and transformed chloroplast genomes, even upon strong overexposure of the blot (data not shown), no signal for the wild-type plastid genome could be detected in the homoplasmic line, indicating that, in three successive selection cycles, all wild-type plastid DNA molecules were successfully eliminated. Successful plastid transformation and homoplasmy was further confirmed by a second RFLP analysis (*Pst*I/*Dra*I; data not shown). In addition, homoplasmy was verified by uniparentally maternal inheritance of the plastid transgene (Fig. 4) and by PCR assays (Fig. 6A).

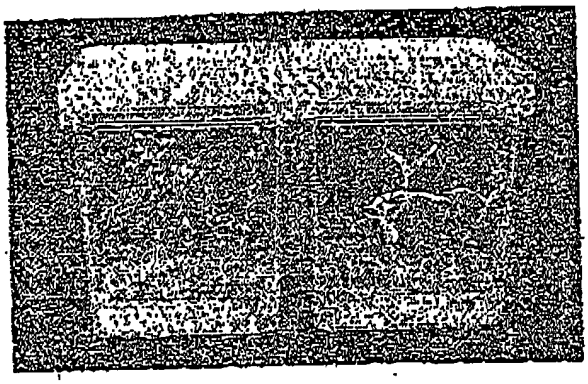


Figure 4. Example of a test for maternal inheritance of the spectinomycin resistance trait in the F1 progeny of transplastomic tomato plants. Flowers from transplastomic plants were pollinated with pollen from wild-type plants, and the seeds were germinated on MS medium containing spectinomycin. Whereas the wild-type control is clearly spectinomycin-sensitive and all seedlings bleach out (right), F1 seedlings from the cross of a transplastomic tomato plant with a wild-type plant exhibit uniform resistance to the antibiotic (left).

even >45% (ref. 10). First we investigated whether the protein accumulation levels in transplastomic tomato plants would be similar to those in transplastomic tobacco. We therefore comparatively analyzed foreign protein accumulation levels in transplastomic tomato plants and in tobacco plants carrying the identical chloroplast transgene (Fig. 5). That no significant difference was found indicated that plastid transgenes in tomato are expressed to similarly high levels as those in tobacco. As expected, transgene expression is stable and foreign protein accumulation levels in the F1 generation are identical to those in the T0 transplastomic generation (Fig. 6B).

Transgene expression levels in the consumable parts of the plant (which frequently are nongreen) are central to the wide use of plastid transformation technology in biotechnology. Most endogenous plastid genome-encoded genes are involved in photosynthesis and hence are drastically downregulated in nonphotosynthetic tissues. Ripe tomato fruits contain chromoplasts, a carotenoid-accumulating differentiation type of plastid. Chloroplasts are present in green tomatoes and are then converted into chromoplasts during the fruit-ripening process. Chromoplasts in tomato fruit were shown to carry out active protein biosynthesis[22] and to contain large amounts of plastid ribosomal RNAs[24]. For this reason, we used an rRNA operon derived chimeric promoter[12] to drive transgene expression in tomato plastid transformation. Additional evidence for the suitability of the rRNA operon promoter for transgene expression in nongreen tissue was provided by the finding that transplastomic tobacco plants carrying a chimeric green fluorescent protein (GFP) gene driven by this promoter exhibit GFP fluorescence not only in green leaves but also in roots, trichomes, and petal chromoplasts[25].

Using an AadA protein–specific antibody, we compared foreign protein accumulation in leaves, green fruits, and ripe red tomatoes. Unexpectedly, high expression levels (approximately half of the expression levels in green leaves) are achieved even in red tomatoes (Fig. 5). Also, foreign protein accumulation does not change significantly during the ripening process.

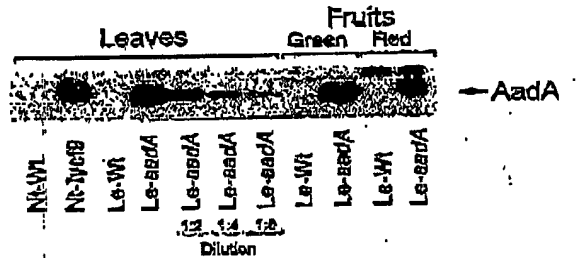


Figure 5. Foreign protein accumulation in leaves, green and ripe red fruits of transplastomic tomato plants. Samples representing 15 µg of extracted total cellular proteins were separated by polyacrylamide gel electrophoresis, blotted to PVDF membranes, and incubated with an AadA-specific polyclonal antibody. As controls, wild-type tomato samples were included for all tissues. For comparison, a transplastomic tobacco line harboring the identical chimeric *aadA* gene (Nt-lycB; ref. 35), as well as a dilution series of leaf proteins from a transplastomic tomato plant, are shown. Note that minor cross-reacting bands in fruit protein extracts are common to wild-type and transplastomic plants and hence do not represent AadA protein.

Our finding that high levels of foreign protein accumulation in transplastomic tomato plants are not confined to photosynthetically active tissue but also occur in ripe fruit opens up new applications for transplastomic technologies, such as the efficient production of nutraceuticals and biopharmaceuticals in plants.

Discussion

Transplastomic technologies may offer a tremendous potential for the production of more nutritious foods. With the successful development of a plastid transformation protocol for tomato, we now have a system at hand that (1) allows plastid genome engineering in a crop for human consumption, (2) results in production of fertile transplastomic plants, and (3) yields high-level foreign protein expression in consumable plant organs. Although earlier work with *Arabidopsis*[17] and potato[17] has resulted in successful chloroplast transformation, all transplastomic *Arabidopsis* lines were male and female sterile and could only be propagated vegetatively in tissue culture[17]. Similarly, the generation of fertile transplastomic potato plants and uniparental transgene transmission to the next generation has not yet been reported (but may be less important for potato than for tomato because potato is largely propagated vegetatively by tuber)[16].

In transplastomic potato plants, foreign protein accumulation was found to be 100-fold lower in nonphotosynthetic microtubers than in green leaves[17]. At present, we can only speculate why transgene expression is so much higher in tomato fruits than in potato microtubers. One reasonable explanation could be that the tomato fruit consisted of green, photosynthetically active tissue before initiation of the ripening process and that active plastid gene expression is maintained upon conversion of chloroplasts to chromoplasts during fruit ripening. In fact, tomato chromoplasts are known to carry out active protein biosynthesis[22] and contain large amounts of plastid rRNAs (ref. 24), whereas the mRNA levels of most photosynthesis-related genes were found to be drastically downregulated. In this respect, the choice of chimeric rRNA operon-derived promoters to drive transgene expression in the tomato fruit may be important for achieving high expression levels. In addition to the selection of the promoter driving transgene transcription, the choice of the translation control signals is a second major determinant for plastid transgene expression levels[26,27]. The translation initiation signals chosen to drive the *aadA* gene in our constructs are not optimal for maximum transgene expression and have been shown to result in 1% foreign protein accumulation (when the transgene was present in a single-copy region of the plastid genome[14]) and 3–5% protein accumulation (when the transgene was present in the inverted-repeat region of the plastid genome[9]). There is likely room for further improvement: combination of the rRNA operon promoter with the bacterio-

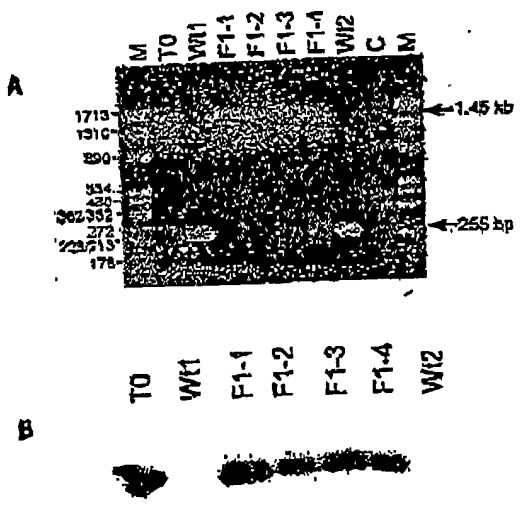


Figure 6. Molecular analyses of F1 progeny resulting from a cross of a transplastomic plant with a wild-type plant. (A) PCR analysis with a primer pair flanking the transgene insertion site in the tomato chloroplast genome confirms transgene presence in the F1 generation in a homoplasmic state. From wild-type plastid genomes, a 255 bp product is amplified (lanes Wt1 and Wt2), whereas from transplastomic genomes a 1.45 kb product is amplified (255 bp chloroplast sequence plus 1.2 kb sequence of the chimeric *aadA* gene). The PCR assay tests for homoplasmy with high sensitivity because it strongly favors amplification of wild-type genome copies because of their smaller size[19]. Complete absence of the 255 bp PCR product and exclusive presence of the 1.45 kb product in both the T0 generation and the F1 progeny (lanes F1-1, F1-2, F1-3, and F1-4 representing four individual F1 plants) confirms stable inheritance of the plastid transgene and successful transformation of all cellular copies of the plastid genome (homoplasmy). M, molecular weight marker (sizes in base pairs indicated at the left); C, buffer control. (B) Western blot analysis confirms stable expression of the plastid transgene in the next generation. Total cellular proteins extracted from leaves were analyzed as in Figure 5.

phage gene 10 leader and a sequence encoding the first 14 amino acids of the GFP yielded a dramatic increase in transgene expression in transplastomic tobacco[25]. Moreover, recent experiments with *Bacillus thuringiensis* (*Bt*) toxin expression in tobacco have demonstrated that foreign protein accumulation in leaves of transplastomic plants can reach values of >45% of the total soluble protein of the cell[10]. Extrapolating from these data, one might expect that at least 20% foreign protein accumulation can theoretically be reached in fruit of transplastomic tomato plants.

Clearly, plastid transformation in tomato is currently much more laborious and time-consuming than plastid transformation in tobacco and *Chlamydomonas*. Our first successful plastid transformation experiments with tomato took almost two years from the biolistic bombardment of the leaves until the harvest of the first ripe transplastomic tomatoes. Certainly, the procedure can be further optimized, and our present efforts aim at making the plastid transformation protocol for tomato more efficient and less time-consuming. The tomato cultivar used in this study (*L. esculentum* var. IAC-Santa Clara) is a red-fruited tomato variety that is commercially grown in South America (average yield: 80–100 t/Ha; fruit weight: 140–170 g). Compared with standard laboratory varieties of tomato, the IAC-Santa Clara displays slower growth in tissue culture and responds significantly less efficiently to shoot induction. Hence, the use of laboratory cultivars potentially could speed up the procedure for generating transplastomic tomato plants, and systematic tests with a number of tomato varieties are currently underway.

Taking advantage of the potentially high transgene expression levels in tomato fruit chromoplasts, this system can now be used to introduce new agronomically and biotechnologically relevant traits into tomato plants by plastid transformation. Experiments are underway that introduce a first set of passenger genes along with the *aadA* selectable marker gene into the tomato plastid genome. Preliminary data indicate successful plastid transformation and stable integration of these passenger genes into the plastid DNA.

The availability of a technology for transgene expression from the tomato plastid genome will open up new possibilities for metabolic engineering, resistance management, and the use of plants as factories for biopharmaceuticals. Plants have considerable potential for the production of edible vaccines, antibodies ("plantibodies"), and therapeutic substances (for a recent review see, e.g., refs 28, 29). For such applications, plastid transformation technologies offer solutions to the technical and ecological problems associated with conventional transgenic technologies (such as transgene silencing and outcrossing[6]) and also achieve high transgene expression levels[9,10,25].

Experimental protocol

Construction of plastid transformation vectors. The *rps12/trnfM* region of plastid transformation vectors. The *rps12/trnfM* region Construction of plastid transformation vectors. The *rps12/trnfM* region (Fig. 1A) was cloned from the tobacco plastid genome as a 3.4 kb *PstI/SbfI* fragment into pBluescript (*PstI/SstI* 3611). A chimeric *aadA* gene driven by the rRNA operon promoter and a synthetic *rbcL*-derived Shine–Dalgarno sequence[12] were inserted into the unique *SpeI* site in between the genes for tRNAGly and tRNAfMet after blunting the *SpeI* site by a fill-in reaction with Klenow enzyme. For subsequent manipulations, a clone was selected in which the *aadA* cassette has the same orientation as the upstream *trnfM* gene (plasmid pRB70; Fig. 1B). pRB70 has the same *aadA* insertion site as the previously described vector pRB15 (ref. 30), but carries a smaller plastid targeting fragment. The remaining polylinker in pRB70 was eliminated by digestion of pRB70 with *ApaI* and *PstI* followed by blunting with mung bean nuclease and re-ligation (clone pRB83). Remaining restriction sites for polylinker enzymes within the *aadA* cassette (*KpnI*, *SpeI*, *XhoI*) were removed by mutagenesis, fill-in reactions with Klenow enzyme, or mung bean nuclease treatment. The complete polylinker from pBSII SK was PCR amplified with M13 primer and reverse primer and the PCR product was cloned into a *HindI* site in between *trnfM* and *aadA*. Clones were selected for both polylinker orientations and control sequenced (plasmids pRB94 and pRB95; Fig. 1C).

Plant material. Sterile tomato plants (*L. esculentum* var. IAC-Santa Clara) were raised in Magenta boxes (double boxes with a connector element) from surface-sterilized seeds germinated on MS medium[31]. For biolistic bombardment, young leaves were harvested from three- to four week-old plants (~15 cm high) produced from outgrowing axillary meristems in stem cuttings. Homoplasmic transplastomic plants and wild-type control plants were transferred to the soil and grown to maturity in a phytochamber (16 h light, 8 h dark, 24°C). Control tobacco plants were grown under identical conditions.

Transformation and regeneration of homoplasmic transplastomic tomato plants. Plastid transformation of tomato was achieved by biolistic bombardment of young sterile tomato leaves with plasmid DNA-coated gold particles of 0.6 μm diameter using the DuPont PDS1000/He biolistic gun[32] and 1,100 p.s.i. rupture disks (BioRad Laboratories, Hercules, CA). Bombarded leaf samples were cut into small pieces (3 × 3 mm), transferred to RMOP medium containing spectinomycin (300–500 mg/L)[4,19], and incubated under dim light (25 μE 16 h light, 8 h dark) for three to four months. Primary spectinomycin-resistant lines were identified as yellow or pale green growing calli. Callus pieces were transferred to the same medium for further propagation and isolation of homoplasmic transplastomic tissue. For plant regeneration, homoplasmic callus tissue was transferred onto the surface of agar-solidified MS medium containing 0.2 mg/L IAA and 3 mg/L BAP. Alternatively, shoot induction was obtained with the same medium but 2 mg/L zeatin instead of BAP. For rooting, regenerated shoots were transferred into boxes containing phytohormone-free MS medium.

DNA extraction, PCR and RFLP analyses. Total cellular DNA was extracted using a cetyltrimethylammonium bromide (CTAB)-based method[33]. For

RFLP analysis, DNA samples were digested with either *PstI* and *XhoI* or *PstI* and *KpnI* (New England Biolabs, Beverly, MA), electrophoresed in 1% agarose gels, and blotted onto nylon membranes (Hybond N, Amersham/Pharmacia, Freiburg, Germany). RFLPs and homoplasmy of transplastomic plants were detected by subsequent hybridization to a radiolabeled *StyI/PstI* restriction fragment (Fig. 1) in RapidHyb buffer following the instructions of the supplier (Amersham/Pharmacia). PCR reactions were performed according to standard protocols (45 s at 94°C, 1.5 min at 55°C, 1.5 min at 72°C; 30 cycles) using primer pairs specific for the chimeric *aadA* gene (primer pair P10/P11 or P28/P29): P10: 5'-AACCTCCTATAGACTAGGC-3'; P11: 5'-AGCGAAATG-TAGTCCTTACG 3'; P28: 5'-TAGCACCCTCTTGATAGAAC-3'; P29: 5'-CGCTATGGAACTCGCCGCC-3'. PCR reactions to assay for homoplasmy employed a primer pair flanking the transgene insertion site in the tomato chloroplast genome (P5'RB70: 5'-GGATTTGGTATAGTTGGCC-3'; P3'RB70: 5'-CTTGTTTATCTATTACTTTTCAGT-3').

Protein extraction and immunoblot analyses. Total soluble protein was extracted from samples homogenized in a buffer containing 300 mM sucrose, 50 mM Tris HCl (pH 8.0), 10 mM ethylene diamine tetraacetic acid, 2 mM ethylene glycol-*O,O'*-bis(2-aminoethyl)-*N,N,N',N'*-tetraacetic acid, 10 mM dithiothreitol, and 1 mM Pefabloc (using a mortar and pestle washed and calcined fine granular quartz; Merck, Darmstadt, Germany). The homogenates were filtered through two layers of Miracloth (Calbiochem, La Jolla, CA) followed by centrifugation at 12,000 r.p.m. for 5 min. For each tissue, samples representing 15 µg of extracted proteins were separated by polyacrylamide gel electrophoresis and blotted to polyvinylidine fluoride (PVDF) membranes. Membranes were subsequently incubated with an AadA-specific polyclonal antibody (raised in rabbits and provided by Dr. Jean-David Rochaix) and detection was carried out with the Western Blot Chemiluminescence Reagent Plus system (NEN Life Science Products, Boston, MA).

Crosses and tests of maternal transgene inheritance. Maternal inheritance of the spectinomycin resistance trait in the F1 progeny of transplastomic tomato plants was analyzed by crosses of transplastomic plants with wild-type plants. Surface-sterilized F1 seeds were germinated on MS medium containing 100 mg/L spectinomycin. Transmission of the *aadA* resistance gene was monitored by the green seedling phenotype and continued growth and development in the presence of the antibiotic in contrast to bleaching and ceased growth of sensitive progeny.

Acknowledgments

We thank Dr. Walter J. Siqueira (Centro de Genética, IAC-Instituto Agronomico de Campinas, Brazil) for tomato seeds. We are grateful to Dr. Jean-David Rochaix (University of Geneva, Switzerland) for providing us with an AadA antibody and Dr. Pal Maliga (Rutgers University, New Jersey) for a chimeric aadA gene. We thank Dr. Jeff Staub (Monsanto Co., St. Louis, MO) for communicating results before publication. This research was supported by a grant in the PROBRAL Program of the Deutscher Akademischer Austauschdienst (DAAD) and the Coordenação Aperfeiçoamento de Pessoal de Nível Superior (CAPES) to H.C. and R.B.

Received 3 April 2001; accepted 22 June 2001

1. Sugiura, M. The chloroplast genome. *Plant Mol. Biol.* 19, 149–168 (1992).
2. Bendich, A.J. Why do chloroplasts and mitochondria contain so many copies of their genome. *BioEssays* 6, 279–282 (1987).
3. Boynton, J.E. *et al.* Chloroplast transformation in *Chlamydomonas* with high-velocity microprojectiles. *Science* 240, 1534–1538 (1988).
4. Svab, Z., Hajdukiewicz, P. & Maliga, P. Stable transformation of plastids in higher plants. *Proc. Natl. Acad. Sci. USA* 87, 8526–8530 (1990).
5. Daniell, H. GM crops: public perception and scientific solutions. *Trends Plant Sci.* 4, 487–489 (1999).
6. Hagel, M. & Beck, R. Enclosed bacteria as new hope for plant biotechnologists. *Appl. Microbiol. Biotechnol.* 54, 302–310 (2000).
7. Heifetz, P.B. Genetic engineering of the chloroplast. *Biochimie* 82, 655–666 (2000).
8. McBride, K.E. *et al.* Amplification of a chimeric *Bacillus* gene in chloroplasts leads to an extraordinary level of an insecticidal protein in tobacco. *BioTechnology* 13, 362–365 (1995).
9. Staub, J.M. *et al.* High yield production of a human therapeutic protein in tobacco chloroplasts. *Nat. Biotechnol.* 18, 333–338 (2000).
10. De Cosa, B., Moar, W., Lee, S.-B., Miller, M. & Daniell, H. Overexpression of the *Bt cry2Aa2* operon in chloroplasts leads to formation of insecticidal crystals. *Nat. Biotechnol.* 19, 71–74 (2001).
11. Staub, J.M. & Maliga, P. Expression of a chimeric *uidA* gene indicates that polycistronic mRNAs are efficiently translated in tobacco plastids. *Plant J.* 7, 845–848 (1995).
12. Staub, J.M. & Maliga, P. Long regions of homologous DNA are incorporated into the tobacco plastid genome by transformation. *Plant Cell* 4, 39–45 (1992).
13. Svab, Z. & Maliga, P. High-frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. *Proc. Natl. Acad. Sci. USA* 90, 913–917 (1993).
14. Carrer, H., Hockenberry, T.N., Svab, Z. & Maliga, P. Kanamycin resistance as a selectable marker for plastid transformation in tobacco. *Mol. Gen. Genet.* 241, 49–56 (1993).
15. Daniell, H., Datta, R., Varma, S., Gray, S. & Lee, S.-B. Containment of herbicide resistance through genetic engineering of the chloroplast genome. *Nat. Biotechnol.* 16, 345–348 (1998).
16. Sikdar, S.R., Serres, G., Chaudhuri, S. & Maliga, P. Plastid transformation in *Arabidopsis thaliana*. *Plant Cell Rep.* 18, 20–24 (1998).
17. Sidorov, V.A. *et al.* Stable chloroplast transformation in potato: use of green fluorescent protein as a plastid marker. *Plant J.* 19, 209–216 (1999).
18. Bogorad, L. Engineering chloroplasts: an alternative site for foreign genes, proteins, reactions and products. *Trends Biotechnol.* 18, 257–263 (2000).
19. Bock, R., Hermann, M. & Kössel, H. *In vivo* dissection of cis-acting determinants for plastid RNA editing. *EMBO J.* 15, 5052–5059 (1996).
20. Hiratsuka, J. *et al.* The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of cereals. *Mol. Gen. Genet.* 217, 185–194 (1989).
21. Maier, R.M., Neckermann, K., Igloi, G.L. & Kössel, H. Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J. Mol. Biol.* 251, 614–628 (1995).
22. Hirose, T., Kusumegi, T., Tsudzuki, T. & Sugiura, M. RNA editing sites in tobacco chloroplast transcripts: editing as a possible regulator of chloroplast RNA polymerase activity. *Mol. Gen. Genet.* 252, 462–467 (1996).
23. Bathgate, B., Purton, M.E., Grierson, D. & Goodenough, P.W. Plastid changes during the conversion of chloroplasts to chromoplasts in ripening tomatoes. *Planta* 164, 197–204 (1985).
24. Piechulla, B., Chamberlin Imlay, K.R. & Gruissem, W. Plastid gene expression during fruit ripening in tomato. *Plant Mol. Biol.* 5, 373–384 (1985).
25. Khan, M.S. & Maliga, P. Fluorescent antibiotic resistance marker for tracking plastid transmission in higher plants. *Nat. Biotechnol.* 17, 910–915 (1999).
26. Ye, G.-N. *et al.* Plastid-expressed 5-enolpyruvylshikimate-3-phosphate synthase genes provide high level glyphosate tolerance in tobacco. *Plant J.* 25, 261–270 (2001).
27. Kuroda, H. & Maliga, P. Sequences downstream of the translation initiation codon are important determinants of translation efficiency in chloroplasts. *Plant Physiol.* 125, 430–436 (2001).
28. Giddings, G., Allison, G., Brooks, D. & Carter, A. Transgenic plants as factories for biopharmaceuticals. *Nat. Biotechnol.* 18, 1151–1155 (2000).
29. Daniell, H., Streatfield, S.J. & Wycoff, K. Medical molecular farming: production of antibodies, biopharmaceuticals and edible vaccines in plants. *Trends Plant Sci.* 6, 219–226 (2001).
30. Bock, R. & Maliga, P. *In vivo* testing of a tobacco plastid DNA segment for guide RNA function in psbL editing. *Mol. Gen. Genet.* 247, 439–443 (1995).
31. Murashige, T. & Skoog, F. A revised medium for rapid growth and bioassays with tobacco tissue culture. *Physiol. Plant.* 15, 493–497 (1962).
32. Kanevski, I. & Maliga, P. Relocation of the plastid *rbcL* gene to the nucleus yields functional ribulose-1,5-bisphosphate carboxylase in tobacco chloroplasts. *Proc. Natl. Acad. Sci. USA* 91, 1969–1973 (1994).
33. Doyle, J.J. & Doyle, J.L. Isolation of plant DNA from fresh tissue. *Focus* 12, 13–15 (1990).
34. Wakasugi, T., Sugita, M., Tsudzuki, T. & Sugiura, M. Updated gene map of tobacco chloroplast DNA. *Plant Mol. Biol. Rep.* 16, 231–241 (1998).
35. Ruf, S., Biehler, K. & Bock, R. A small chloroplast-encoded protein as a novel architectural component of the light-harvesting antenna. *J. Cell Biol.* 149, 369–377 (2000).
36. Svab, Z. & Maliga, P. Mutation proximal to the tRNA binding region of the *Nicotiana* plastid 16S rRNA confers resistance to spectinomycin. *Mol. Gen. Genet.* 228, 316–319 (1991).

HERMANN *ET AL.* 1999

Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 4856–4861, April 1999
Biochemistry

Transfer of plastid RNA-editing activity to novel sites suggests a critical role for spacing in editing-site recognition

MARITA HERMANN AND RALPH BOCK*

Institut für Biologie III, Universität Freiburg, Schänzlestrasse 1, D-79104 Freiburg, Germany

Edited by C. S. Levings III, North Carolina State University, Raleigh, NC, and approved February 23, 1999 (received for review December 21, 1998)

ABSTRACT **RNA editing in higher plant plastids alters mRNA sequences by C-to-U conversions at highly specific sites through an unknown mechanism. To elucidate how the cytidine residues to be edited are specifically recognized and distinguished from other cytidines in close proximity, we have changed *in vivo* the distances of two plastid RNA-editing sites from their essential upstream cis-acting sequence element. Analysis of RNA editing in transgenic chloroplasts revealed that reduction of this distance by 1 nt entirely abolishes RNA editing. Surprisingly, deletions or combinations of deletional and point mutations that shift a heterologous cytidine residue in the same distance from the upstream cis-element as the editing site in the wild type result in transfer of the RNA-editing activity to the heterologous cytidine whereas the wild-type site remains unedited. Our results suggest that the molecular identity of at least some editing sites in the chloroplast genome is defined by their distance from an essential upstream sequence element.**

Posttranscriptional alterations of single nucleotides within an mRNA are referred to as RNA editing and have been described for a variety of genetic systems (1), including higher-plant mitochondria (2–4) and chloroplasts (refs. 5 and 6; for review, see, e.g., ref. 7). It appears useful to formally distinguish between two major types of RNA editing: insertional/deletional and conversional editing. The insertional/deletional type of editing is best known from kinetoplasts of trypanosomes, where uridine residues are inserted or deleted (ref. 8; for review, see, e.g., ref. 9). Editing in mammalian nuclei and plant organelles is of the conversional type and typically involves purine-to-purine or pyrimidine-to-pyrimidine transitions. The editing processes in different genetic systems employ widely different mechanisms, implying that editing activities may have evolved several times independently.

The editing systems in higher-plant mitochondria and chloroplasts share many similar features and, thus, may have originated from common evolutionary roots. RNA editing in both organelles proceeds mainly by C-to-U conversions with the exception of few reverse events. Editing is an early posttranscriptional event and an essential processing step in the maturation of organellar transcripts: the nucleotide conversions usually alter the coding properties of the mRNA, thereby facilitating the synthesis of functional proteins (10).

A central question surrounding plant organellar RNA editing is how to explain the extraordinarily high specificity with which the editing apparatus selects individual cytosine residues for modification. The sequences flanking editing sites lack any apparent conserved, consensus sequence-like elements at the primary or at the secondary structure level. A number of *in vivo* studies in transgenic chloroplasts have demonstrated that mRNA sequences flanking the editing site are involved directly in plastid RNA editing (11–14). However, the molecular mechanism by which RNA-editing sites are recognized with high specificity as well as how the editing machinery distinguishes between the cytidines to be edited and other cytidines in close proximity are completely unknown. The absence of consensus motifs at the mRNA level may indicate that editing sites in plant organelles are selected by a molecular mechanism different from the recognition of primary or secondary structural features at the editing site itself.

Here we provide evidence from transgenic *in vivo* studies that the position of a cytidine residue in relation to an essential upstream cis-acting sequence element determines whether or not this cytidine can be edited. Our results suggest a model for editing site recognition in which the editing machinery binds to an upstream sequence element and recognizes the editing site as being a downstream cytidine in a defined distance.

MATERIALS AND METHODS

Plant Material. Tobacco plants (*Nicotiana tabacum* cv Petit Havana) were grown under sterile conditions on agar-solidified MS medium (15) containing 30 g/liter sucrose. Homoplasmic transformed lines were rooted and propagated on the same medium. The previously generated transplastomic tobacco line Nt-pRB59 (12) was kept under identical conditions. For seed assays and tests of maternal transgene inheritance, wild-type and transformed plants were transferred to soil and grown to maturity under standard greenhouse conditions.

List of Oligonucleotides. Oligonucleotides included: P3, 5′-CAGTTGGAAGAATTTGTCC-3′; P10, 5′-AACCTCCTATAGACTAGGC-3′; P11, 5′-AGCGAAATGTAGTGCTTACG-3′; P16, 5′-TTTTTCTAGACGCTCATATTCATTACCGTA-3′; P28, 5′-TAGCACCCTCTTGATAGAAC-3′; P29, 5′-CGCTATGGAACTCGCCGCC-3′; P30, 5′-TTTTGGATCCTACGTCAGGAGTCCATTGATGAGAAGGGCTGGGGA-3′; P31, 5′-TTTTGGATCCTACGTCAGGAGTCCATTGATGAGAAGGCTGGGGA-3′; P32, 5′-TTTTGGATCCTACGTCAGGAGTCCATTGATGAGAAGGGGCTGGGGGAAAGC-3′; P33, 5′-TTTTGGATCCTACGTCAGGAGTCCATTGATAGGAAGGGCTGGGGA-3′; 7355, 5′-GACTATAGATCGAACCTATCC-3′; 1020, 5′-CAAGATCCATTACGTGTCCAAGG-3′.

Construction of Plastid Transformation Vectors. Chimeric genes containing mutated sequences with the *ndhB* editing sites IV and V were constructed by using the previously generated plastid transformation vector pRB51, which contains a minilinker between the *aadA* coding region and the *psbA* 3′ untranslated region (12). Editing sites IV and V-containing *ndhB* fragments for insertion into pRB51 were prepared by PCR amplification of the corresponding plastid sequences (16) from −42, with respect to editing site IV, to +22, with respect to editing site V. A 5′ *Xba*I restriction site,

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

This paper was submitted directly (Track II) to the *Proceedings* office.
Abbreviation: cm, compensatory mutation.
*To whom reprint requests should be addressed. e-mail: bock@biologie.uni-freiburg.de.

a 3′ *Bam*HI site, and the desired (insertional or deletional) mutations were introduced with the primer sequences. After digestion with *Xba*I and *Bam*HI, the fragments were cloned into the similarly cut pRB51 and the correctness of mutagenesis and cloning were verified by DNA sequencing with primer P3. In this way, the following transformation vectors were generated: pRB67-Δ1 (*ndhB* insertion amplified with the primer pair P16/P30), pRB68-Δ2 (primer pair P16/P31), pRB69-i1 (primer pair P16/P32), and pRB73-Δ1/cm (compensatory mutation; primer pair P16/P33).

Plastid Transformation and Selection of Transplastomic Tobacco Lines. Young leaves were harvested from sterile plants and bombarded with plasmid-coated tungsten particles by using the DuPont PDS1000He biolistic gun (17). Primary spectinomycin-resistant lines were selected on RMOP regeneration medium containing 500 mg/liter spectinomycin dihydrochloride (18). PCR, using the chimeric *aadA* gene-specific primer pair P10/P11, eliminated spontaneous, spectinomycin-resistant mutants and identified true plastid transformants. Correct integration of the constructs into the chloroplast genome was verified by PCR with the primer pair P11/1020 (12). For each construct, three independent, transplastomic lines were subjected to four additional rounds of regeneration on RMOP/spectinomycin to obtain homoplasmic tissue. Homoplasmy was verified by a highly sensitive PCR assay (12).

Isolation of Nucleic Acids. Total plant nucleic acids were extracted according to a rapid miniprep procedure described by Doyle and Doyle (19). Total cellular RNA was isolated by using the TRIzol reagent (GIBCO/BRL). For cDNA synthesis, an aliquot of the RNA preparation was treated with DNase I. Vector DNA for biolistic transformation and templates for plasmid sequencing were prepared by using the Qiagen column-purification system.

cDNA Synthesis and PCR. Reverse transcription was primed with a random hexanucleotide primer mixture for 10 min at room temperature. The elongation reaction was carried out with SuperScript II reverse transcriptase (GIBCO/BRL) at 42°C following the manufacturer's instructions. DNA and cDNA templates were amplified according to standard PCR protocols.

DNA Sequencing. Plasmid DNA was sequenced by cycle sequencing, using the fluorescence-labeled oligonucleotide P3 as sequencing primer. Primer pairs P11/P28 or P29/P28 were used to generate the substrate for direct sequencing of transgene-derived PCR products. Amplification products were purified for sequencing by electrophoresis on 2% agarose gels and subsequent extraction from excised gel slices by using the Qiaex II kit (Qiagen). Sequence determination was carried out by a modified chain-termination method (20). Oligonucleotide P28 served as sequencing primer for the PCR products. RNA-editing efficiencies were quantitated by using a PhosphorImager and a quantitation procedure developed earlier (12).

RESULTS

Integration of Mutated Sequences Containing Two *ndhB* Editing Sites into the Tobacco Plastid Genome. The plastid *ndhB* gene encodes a subunit of a putative chloroplast NADH dehydrogenase (21). The *ndhB* mRNA was shown previously to undergo several base changes by RNA editing (22, 23). It contains nine editing sites in tobacco (23), six of which are grouped in three pairs with two closely spaced sites each.

We recently have defined a minimum sequence context that is necessary and sufficient to direct editing at *ndhB* sites IV and V *in vivo* (12). In these analyses, a sequence element 5′ upstream of the two editing sites (in the −2 to −12 region with respect to site IV) was identified that is absolutely required for eliciting the editing reaction at both sites. Most of the nucleotides in the small, 8-nt spacer in between sites IV and V turned out to have little or no influence on editing efficiencies (14). Both editing sites are embedded in sequences with numerous other cytidine residues in close proximity (Fig. 1), and it is not clear how the plastid RNA-editing machinery distinguishes between these cytidines and recognizes the editing sites with high specificity.

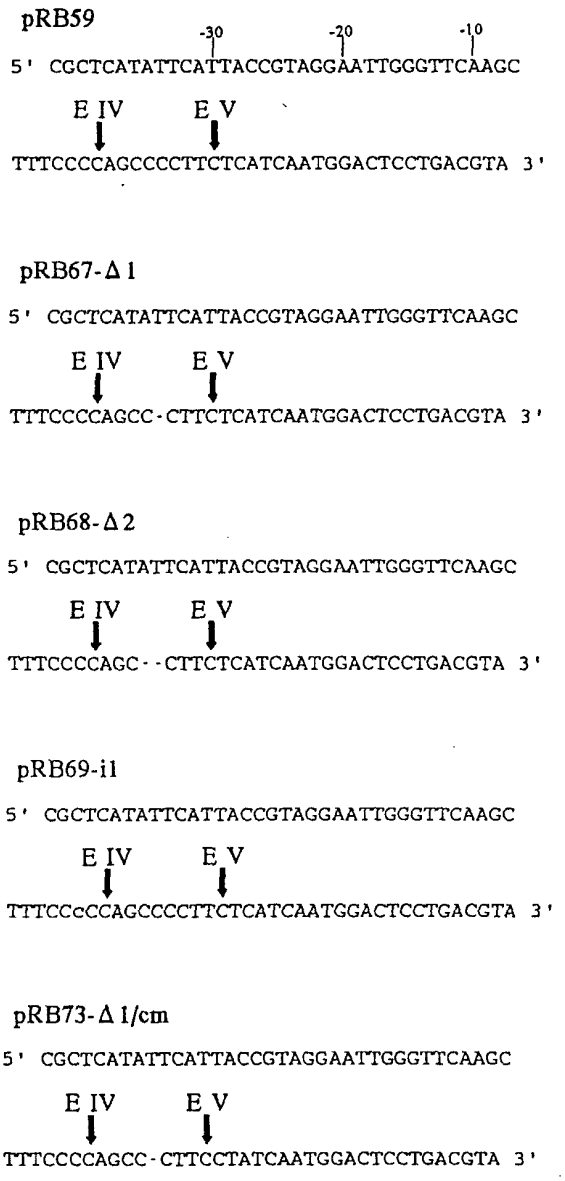


FIG. 1. Sequences of the *ndhB* segment insertions in the chloroplast transformation vectors used in this study. Plasmid pRB59 contains the wild-type sequence (12) and spans an *ndhB* segment from −42, with respect to the 5′ editing site (site IV), to +22, with respect to the 3′ editing site (site V). Vector pRB67-Δ1 carries a deletion of a single nucleotide in the 8-nt spacer separating sites IV and V. pRB68-Δ2 has a deletion of 2 nt in the spacer, and in pRB69-i1, a single C residue was inserted into a tetracytidine motif harboring editing site IV (denoted by a lowercase C at an arbitrarily chosen position). Vector pRB73-Δ1/cm is identical to pRB67-Δ1 but carries an additional compensatory point mutation expected to restore editing at the 3′ site. See text for details. Editing sites as in the wild type are marked by vertical arrows. Deleted nucleotides are indicated by dashes, and inserted nucleotides are shown as lowercase letters. The nucleotides upstream of site IV (harboring the essential cis-acting elements for the editing of both sites in the −2/−12 region; ref. 12) are numbered in the pRB59 sequence, with editing site IV as nucleotide "0."

The requirement for a cis-acting element in proximity to, but spatially separated from, the editing sites raises an attractive possibility of how the cytidines to be edited are specifically recognized: the editing apparatus could act only on those cytidines that are in an exactly defined distance from the essential upstream sequence element. To test this hypothesis, we have constructed a series of chloroplast transformation vectors in which this distance was altered. For this purpose, we chose the *ndhB* editing sites IV and V as a model system because they are well characterized with respect to the sequence requirements for editing.

The *ndhB* sequence manipulations were carried out in a −42/+22 segment spanning the two editing sites IV and V. Incorporated into a chimeric context, this segment was shown to yield 95% editing at site IV and 75% at site V (12, 14). In the final transformation vectors, the mutated *ndhB* sequence segment is linked to and cotranscribed with a selectable marker gene (*aadA*) conferring resistance to spectinomycin (12, 17). The flanking regions of homology to sequences of the tobacco plastid genome target the chimeric *aadA/ndhB* transgenes to the intergenic spacer region between the *psbE* operon and the *petA* gene, which is known to be a suitable target site for the uptake of transgenes (10, 12). In this way, the transformation vectors with the *ndhB* insertions shown in Fig. 1 were constructed.

The chimeric *aadA/ndhB* genes were integrated into the tobacco plastid genome by using the biolistic process. Sterile tobacco leaves were bombarded with plasmid DNA-coated tungsten particles and subsequently subjected to selection for spectinomycin resistance on a plant-regeneration medium (18). Primary plastid transformants appeared after 1–2 months. Correct integration of the transgene into the chloroplast genome was confirmed by PCR-based assays. Subsequently, homoplasmic transplastomic lines were purified by repeated plant regeneration under selective conditions.

Deletion of a Single Nucleotide in the Spacer Between Editing Sites IV and V Selectively Abolishes Editing at Site V. In a first experiment toward determining the influence of nucleotide phasing on editing-site recognition, we left the 5′ editing site (site IV) and its distance from the upstream cis-element unchanged and merely altered the distance of the downstream site V by 1 nt. Analysis of partially edited cDNA clones recently has established that the two editing sites are edited independently and not in a 3′ → 5′ or 5′ → 3′ polar fashion. Consequently, loss of editing at one site is not expected to abolish editing at the other (14). Moreover, using a scanning-point mutagenesis approach, the four cytidine residues in the 8-nt spacer between the two editing sites (Fig. 1) were shown not to be involved in editing-site recognition (14). Thus, this C_4 motif seemed to be a suitable site at which to introduce the desired mutations. We first deleted one of the cytidine residues from the spacer between editing sites IV and V (pRB67-Δ1, Fig. 1) and generated transgenic tobacco plants carrying this mutated *ndhB* segment in their chloroplast genome.

Sequencing of the cDNA population derived from the chimeric *aadA/ndhB* gene construct revealed that editing at the 5′ site (IV) is not affected by the single nucleotide deletion (Fig. 2*A*). This finding is consistent with the idea that the essential elements for editing-site recognition reside upstream of both sites (12). Editing at site V, however, turned out to be completely abolished in the Nt-pRB67-Δ1 transplastomic tobacco lines. Because the nucleotide deleted in pRB67-Δ1 can be changed by point mutagenesis without any effect on editing of site V (13), loss of site V editing in the Nt-pRB67-Δ1 lines could indicate that, indeed, the distance of the editing site from the upstream cis-element is critical for editing-site selection.

Deletion of Two Nucleotides from the Spacer Between Editing Sites IV and V Induces Editing at a Novel Site. We then deleted two of the four cytidines in the spacer between editing sites IV and V (pRB68-Δ2; Figs. 1 and 2*B*). This deletion is different from the above-described mutation in that it shifts a downstream cytidine in place of the wild-type editing site V (Figs. 1 and 2*B*). If the distance from the essential upstream cis-element indeed were the major determinant for editing-site recognition, then the editing machinery would now find a cytidine residue in the correct phase with the cis-element and possibly would be able to act on this heterologous substrate cytidine.

As expected, editing at the 5′ site IV also was not affected by the 2-nt deletion present in the Nt-pRB68-Δ2 plants. Also, as in the Nt-pRB67-Δ1 lines, editing at site V was completely lost in the Nt-pRB68-Δ2 plants. However, the novel "in-phase" cytidine now undergoes editing with efficiency (65%) similar to that in site V in the Nt-pRB59 control lines (75%, Fig. 2). This transfer of the editing activity to a heterologous site may suggest that this editing site indeed may be recognized as a cytidine being in a defined distance from an upstream cis-acting sequence element.

A Single-Nucleotide Insertion Upstream of Site IV Affects Editing at Both Sites. We next wanted to test whether the identity of the upstream editing site IV is determined similarly by its distance from the upstream cis-element, as is shown for site V. Site IV is part of a tetracytidine motif, with the 3′ most cytidine as the editing position (Fig. 1). We inserted an additional cytidine into this motif. If the distance from the 5′ cis-element was also the critical determinant for site IV, then the fourth of the now five cytidines should undergo editing. Alternatively, if the position of the editing site in the surrounding sequence context was the recognition principle, then, as in the wild type, the 3′ most (i.e., the fifth) cytidine should be recognized. Simultaneously, this cytidine insertion changes the distance of editing site V from the upstream cis-element. In contrast to the Nt-pRB67-Δ1 and Nt-pRB68-Δ2 transplastomic lines, where this distance is reduced by 1 or 2 nt, respectively, it now is increased by 1 nt in transformation vector pRB69-i1 (Fig. 1).

Analysis of site IV editing in the Nt-pRB69-i1 transplastomic lines revealed that only the fourth cytidine in the pentacytidine motif was edited (Fig. 3). No editing was detected at the 3′ most cytidine edited in the wild type, suggesting that, for *ndhB* editing site IV as well, the position of the cytidine in relation to the upstream cis-acting element determines the identity of the editing site. In the Nt-pRB69-i1 lines, cytidines are both the 5′ and the 3′ neighboring nucleotides of the edited cytidine. Our finding that, nonetheless, only the cytidine in the correct distance from the upstream cis-element undergoes editing is indicative of a remarkably high accuracy with which this recognition mechanism operates.

The editing efficiency at the fourth cytidine is significantly lower (approximately 20%) than that at site IV in the Nt-pRB59 control lines (95%). This most probably is caused by the presence of a different 3′ neighboring nucleotide of the editing site. This explanation is in accordance with the earlier findings that the nucleotides immediately adjacent to the editing position contribute significantly to the efficiency of the editing reaction (12–14).

As expected from the results with the nucleotide deletions in the spacer region (as in pRB67-Δ1 and pRB68-Δ2), the insertion of the cytidine in the Nt-pRB69-i1 lines also exerts a negative effect on RNA editing at downstream site V. However, we reproducibly detected a residual editing activity of approximately 10% (Fig. 3). Thus, in contrast to the deletion of 1 nt, which leads to a complete loss of editing, insertion of 1 nt does not entirely abolish editing. From an evolutionary point of view, such a slightly relaxed specificity would be tolerable because the nucleotides 5′ and 3′ of editing site V are not cytidines. However, in the case of site IV, which is flanked by other cytidines, the editing machinery seems to measure the distance from the upstream cis-element with perfect accuracy,

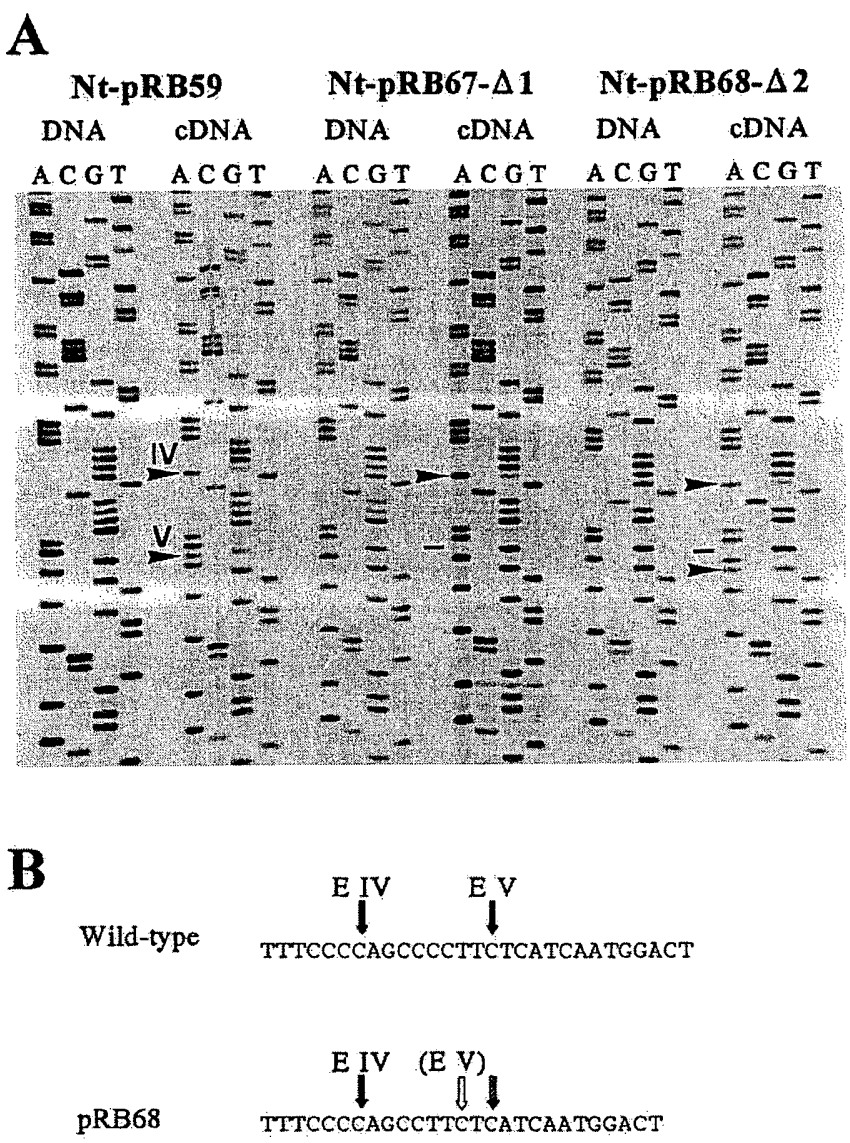


FIG. 2. RNA editing in the Nt-pRB67-Δ1 and Nt-pRB68-Δ2 transplastomic tobacco plants. (*A*) Sequence analysis to test for transgene mRNA editing in the Nt-pRB67-Δ1 and Nt-pRB68-Δ2 transplastomic lines in comparison with editing of the wild-type *ndhB* sequence placed in an identical transgenic context (line Nt-pRB59). DNA and cDNA were amplified with primer pair P11/P28 and sequenced directly with oligonucleotide P28. Owing to the polarity of this primer, the sequence ladders reflect the DNA strand complementary to the mRNA sequence. Arrowheads point to the editing positions in the cDNA lanes (G in DNA; A in cDNA), dashes denote lack of editing, and roman numerals indicate editing sites IV and V (see Fig. 1). Note the lack of editing at site V in Nt-pRB67-Δ1 (dash) and induction of editing at a downstream cytidine (arrowhead) in Nt-pRB68-Δ2. (*B*) Shift of the *ndhB* site V editing activity to a heterologous site in Nt-pRB68-Δ2 transplastomic tobacco lines. Deletion of two cytidine residues from the spacer sequence between editing sites IV and V results in the transfer of a downstream cytidine (stippled arrow) in a distance from site IV (solid arrow) identical to that of what was previously site V (solid arrow in the wild-type sequence; open arrow in the pRB68-Δ2 sequence). Whereas wild-type site V remains unedited, editing is induced at the cytidine, which is now 9 nt from site IV.

thereby preventing misediting of neighboring cytidines, which potentially would result in the synthesis of nonfunctional proteins.

A Compensatory Point Mutation Restores RNA Editing in the Single-Nucleotide Deletion Mutant. As described above, deletion of a single nucleotide in the spacer region between editing sites IV and V led to a complete loss of editing at site V (Nt-pRB67-Δ1 lines, Fig. 2). To provide additional evidence for the distance being the critical determinant for editing-site selection, we have attempted to restore editing in these mutants by the introduction of a compensatory point mutation that creates a cytidine residue in the "correct" distance from the upstream cis-element. Construction of this double mutant was accomplished by changing the T nucleotide, which was in the potentially editable position in vector pRB67-Δ1 (Fig. 1), into a C. The original sequence context (i.e., the 3′ neighboring nucleotide known to influence the efficiency of the editing reaction) was maintained partially by an additional C-to-T change immediately downstream, which restores the 3′ thymidine flanking editing site V in the wild-type sequence (Fig. 1). The resulting transformation vector was termed pRB73-Δ1/cm (Fig. 1). If the distance from the upstream sequence element determines the editing-site selection, then the prediction is that editing now should take place at the newly created cytidine downstream of editing site V (Figs. 1 and 4*B*).

Analysis of the transgene-derived cDNA population from the generated Nt-pRB73-Δ1/cm transplastomic tobacco lines revealed that, as predicted, the introduced compensatory point mutation restores RNA editing in the single-nucleotide deletion mutant (Fig. 4). The editing efficiency (calculated to be

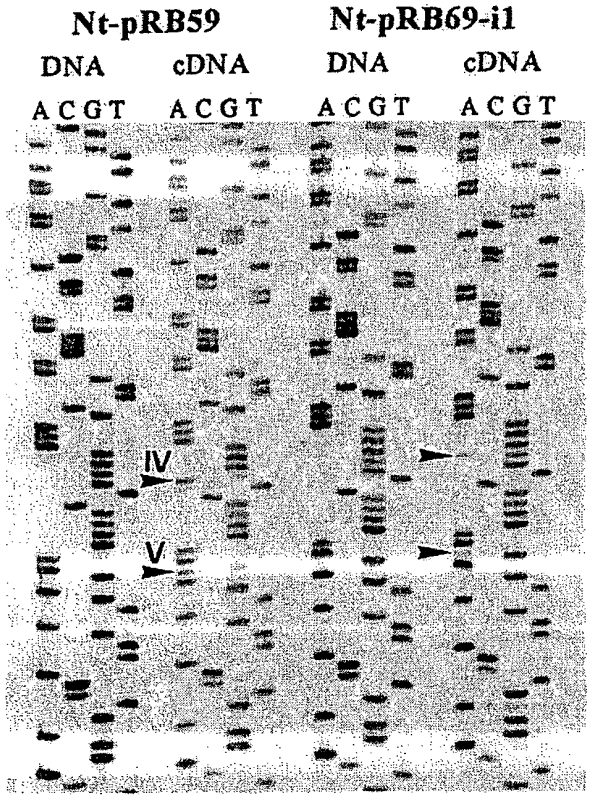


FIG. 3. Sequence analysis to test the effect of a single C nucleotide inserted into a tetracytidine motif containing editing site IV as the 3′ most cytidine residue. For comparison, editing in the wild-type *ndhB* sequence as contained in line Nt-pRB59 also is shown. DNA and cDNA samples were amplified with primer pair P11/P28 and sequenced directly with primer P28. Because of the polarity of this primer, the autoradiograph shows the sequences of the DNA strand complementary to the mRNA. Arrowheads mark the editing positions in the cDNA lanes, and roman numerals indicate editing-site numbers.

approximately 35%) is lower compared with that at site V in the pRB59 control lines (75%). This most probably is due to the presence of a different neighboring nucleotide 5′ of the editing site (T in the wild type and C in the Nt-pRB73-Δ1/cm lines, Fig. 4*B*). This 5′ neighboring position was shown earlier to exert a significant influence on the efficiency of the editing reaction (12–14).

DISCUSSION

Chloroplast transformation, though being a laborious and time-consuming technology, is currently the method of choice for the study of RNA editing in higher-plant plastids. Earlier studies attempted to define the cis-acting sequence requirements for RNA-editing-site selection and to define minimum substrates for the plastid editing machinery (12, 13). Using two well characterized *ndhB* editing sites from tobacco, the scope of this study was to test whether the distance of plastid editing sites from an upstream cis-element could be a determinant for editing-site selection. This could explain how the editing apparatus selectively recognizes the editing site and distinguishes between the cytidine to be edited and other cytidines in the immediate neighborhood.

We report here that the recognition of the two *ndhB* RNA-editing sites is critically dependent on their distance from an upstream essential cis-acting element. Apparently, only those cytidine residues that are the correct distance from this upstream element can be recognized by the editing apparatus. Small changes of this distance can abolish editing

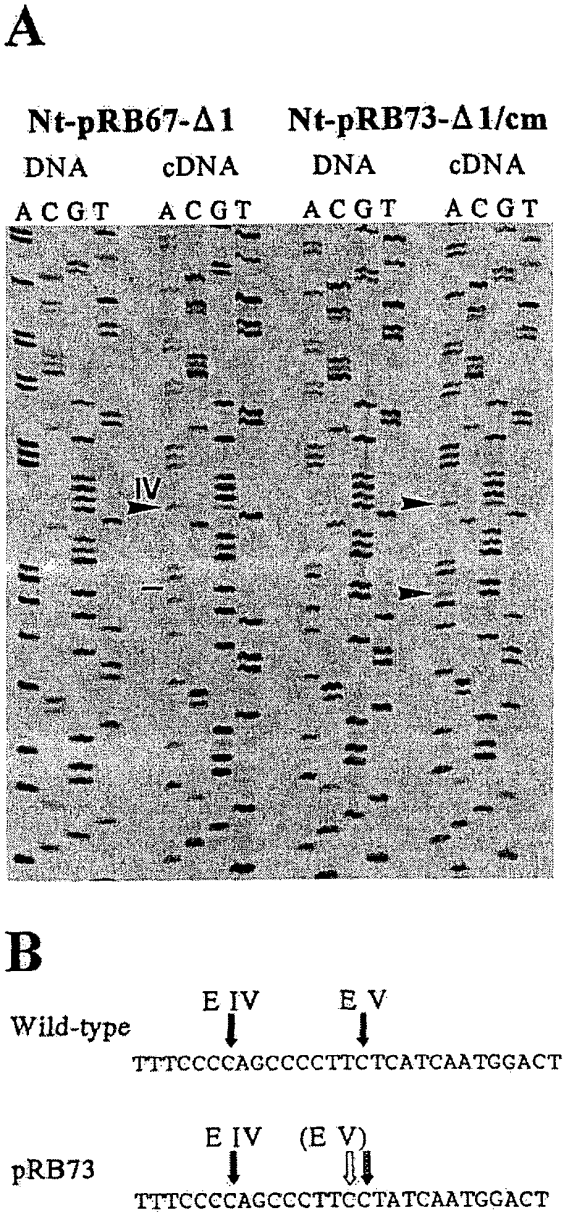


FIG. 4. Restoration of site V editing in the single-nucleotide-deletion mutant (construct pRB67-Δ1) by introduction of a compensatory point mutation, creating an editable C the correct distance from the upstream cis-element (construct pRB73-Δ1/cm). (*A*) Sequence analysis to test for transgene mRNA editing in the Nt-pRB73-Δ1/cm transplastomic lines. For comparison, editing in the Nt-pRB67-Δ1 lines also is shown. PCR and sequencing primers are as in Fig. 3. Arrowheads point to the editing positions in the cDNA lanes (G in DNA; A in cDNA), a dash indicates lack of editing, and roman numerals mark editing site IV. Note the lack of editing at site V in Nt-pRB67-Δ1 (dash) and restoration of editing after conversion of the downstream nucleotide into a cytidine in Nt-pRB73-Δ1/cm (arrowhead). (*B*) Restoration of editing in the Nt-pRB73-Δ1/cm transplastomic lines. The single-nucleotide deletion in the spacer, abolishing editing at site V in Nt-pRB67-Δ1 (solid arrow in the wild-type sequence; open arrow in the pRB73-Δ1/cm sequence), is compensated by mutational creation of an "in-phase" cytidine immediately 3′ of editing site V (stippled arrow).

completely, indicating that this distance is "measured" by the editing apparatus with high accuracy. We have shown that heterologous cytidine residues can undergo editing when placed the correct distance downstream of the cis-element,

suggesting that this distance is a major determinant for selection of the correct cytidine for modification by the editing machinery.

C-to-U editing at a single position in the mammalian apolipoprotein B (apoB) mRNA was demonstrated to involve an essential RNA sequence element in close proximity to the editing site. This cis-acting element was termed the "mooring sequence" and is believed to mediate both substrate recognition and editosome assembly (for review, see, e.g., ref. 24). In this system, the editosome will edit any cytidine that is located in a 3- to 5-nt distance 5′ from the mooring sequence. In addition to the much higher number of editing sites the editing machineries in plant organelles have to deal with, plastid editing—at least in the case of the *ndhB* sites examined here—differs in two aspects from mammalian *apoB* editing. First, the mooring sequence-like, essential cis-acting element resides upstream of the editing site for plastid *ndhB* editing but downstream of the editing site in the case of *apoB* editing. Second, whereas there is a larger window for the distance of the editing site from the mooring sequence in *apoB* editing, this distance seems to be more precisely defined in plastid RNA editing. At present, we can only speculate about a possible evolutionary relationship between mammalian C-to-U editing and the editing systems in plant organelles. Clarification of this point would require a thorough comparison of the factors involved in editing in both systems. However, whereas the editing enzyme for *apoB* editing meanwhile is well characterized, the trans-factors involved in plant organellar RNA editing still await their molecular identification.

There is now compelling evidence for the participation of at least two factors in the editing reactions in plastids: an essential cis-acting element at the mRNA level (12, 13) and a site-specific trans-acting factor of unknown molecular identity (11, 25). By analogy to *apoB* editing, the essential upstream sequence element could serve as mooring sequence, allowing for binding of the editing apparatus to its RNA substrate mediated by the site-specific trans-factor. For *ndhB* editing sites IV and V, this model implies the existence of two specificity factors (one site IV-specific and one site V-specific) accounting for the "measuring" of two distinct distances from the essential upstream cis-element. This upstream sequence element harbors either a single mooring sequence employed by both specificity factors or two distinct, but largely overlapping mooring sequences (evidenced by deletion of this region, which abolishes editing at both sites; ref. 12). The existence of separate specificity factors for sites IV and V is also in agreement with the earlier finding that the two sites are edited independently (14). The data presented here support the idea that plastid editing sites are recognized specifically by a sophisticated interplay of a cis-acting element and a site-specific trans-acting factor. In addition to these qualitative determinants, other factors are known to influence the efficiency of the editing reaction in a quantitative fashion: (*i*) the identity of the nucleotides immediately adjacent to the editing site (13, 14) and (*ii*) upstream as well as downstream sequence elements outside the minimum sequence context (12, 13) that have not yet been characterized in detail.

It remains to be determined whether the presence of an upstream cis-element and the distance of the editing site from it are also the major determinants for the recognition of other plastid or even plant mitochondrial editing sites. Given the laborious and time-consuming procedures involved in the generation of homoplasmic plants with transgenic chloroplasts, a major obstacle in this respect is posed by the lack of efficient *in vitro* systems for plant organellar RNA editing (26, 27). Therefore, the development of faithful *in vitro* assays for plant RNA editing represents one of the major challenges for the future.

We thank Drs. Pal Maliga and Zora Svab (Rutgers University, Piscataway, NJ) for providing the plastid *aadA* gene. We thank Dr. Jörg Kudla (University of Ulm, Germany) for helpful suggestions for the preparation of this manuscript. This research was supported by grants from the Deutsche Forschungsgemeinschaft to R.B. (Bo 1482/3-2).

1. Smith, H. C., Gott, J. M. & Hanson, M. R. (1997) *RNA* **3**, 1105–1123.
2. Covello, P. S. & Gray, M. W. (1989) *Nature (London)* **341**, 662–666.
3. Gualberto, J. M., Lamattina, L., Bonnard, G., Weil, J.-H. & Grienenberger, J. M. (1989) *Nature (London)* **341**, 660–662.
4. Hiesel, R., Wissinger, B., Schuster, W. & Brennicke, A. (1989) *Science* **246**, 1632–1634.
5. Hoch, B., Maier, R. M., Appel, K., Igloi, G. L. & Kössel, H. (1991) *Nature (London)* **353**, 178–180.
6. Kudla, J., Igloi, G. L., Metzlaff, M., Hagemann, R. & Kössel, H. (1992) *EMBO J.* **11**, 1099–1103.
7. Maier, R. M., Zeltz, P., Kössel, H., Bonnard, G., Gualberto, J. M. & Grienenberger, J. M. (1996) *Plant Mol. Biol.* **32**, 343–365.
8. Benne, R., Van den Burg, J., Brakenhoff, J. P. J., Sloof, P., Van Boom, J. H. & Tromp, M. C. (1986) *Cell* **46**, 819–826.
9. Kable, M. L., Heidmann, S. & Stuart, K. D. (1997) *Trends Biochem. Sci.* **22**, 162–166.
10. Bock, R., Kössel, H. & Maliga, P. (1994) *EMBO J.* **13**, 4623–4628.
11. Chaudhuri, S., Carrer, H. & Maliga, P. (1995) *EMBO J.* **14**, 2951–2957.
12. Bock, R., Hermann, M. & Kössel, H. (1996) *EMBO J.* **15**, 5052–5059.
13. Chaudhuri, S. & Maliga, P. (1996) *EMBO J.* **15**, 5958–5964.
14. Bock, R., Hermann, M. & Fuchs, M. (1997) *RNA* **3**, 1194–1200.
15. Murashige, T. & Skoog, F. (1962) *Physiol. Plant.* **15**, 473–497.
16. Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., *et al* (1986) *EMBO J.* **5**, 2043–2049.
17. Svab, Z. & Maliga, P. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 913–917.
18. Svab, Z., Hajdukiewicz, P. & Malgia, P. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8526–8530.
19. Doyle, J. J. & Doyle, J. L. (1990) *Focus* **12**, 13–15.
20. Bachmann, B., Lüke, W. & Hunsmann, G. (1990) *Nucleic Acids Res.* **18**, 1309.
21. Burrows, P. A., Sazanov, L. A., Svab, Z., Maliga, P. & Nixon, P. J. (1998) *EMBO J.* **17**, 868–876.
22. Maier, R. M., Neckermann, K., Hoch, B., Akhmedov, N. B. & Kössel, H. (1992) *Nucleic Acids Res.* **20**, 6189–6194.
23. Freyer, R., López, C., Maier, R. M., Martin, M., Sabater, B. & Kössel, H. (1995) *Plant Mol. Biol.* **29**, 679–684.
24. Chan, L., Chang, B. H.-J., Nakamuta, M., Li, W.-H. & Smith, L. C. (1997) *Biochim. Biophys. Acta* **1345**, 11–26.
25. Bock, R. & Koop, H.-U. (1997) *EMBO J.* **16**, 3282–3288.
26. Araya, A., Domec, C., Begu, D. & Litvak, S. (1991) *Proc. Natl. Acad. Sci. USA* **89**, 1040–1044.
27. Yu, W. & Schuster, W. (1995) *J. Biol. Chem.* **31**, 18227–18233.

MAIER *ET AL.* 1995

J. Mol. Biol. (1995) 251, 614–628

JMB

Complete Sequence of the Maize Chloroplast Genome: Gene Content, Hotspots of Divergence and Fine Tuning of Genetic Information by Transcript Editing

Rainer M. Maier, Kai Neckermann, Gabor L. Igloi and Hans Kössel*

Institut für Biologie III der Universität Freiburg Schänzlestr. 1, D-79104 Freiburg, Germany

The nucleotide sequence of the chloroplast (cp) DNA from maize (*Zea mays*) has been completed. The circular double-stranded DNA, which consists of 140,387 base-pairs, contains a pair of inverted repeat regions (IR_A and IR_B) with 22,748 base-pairs each, which are separated by a small and a large single copy region (SSC and LSC) of 12,536 and 82,355 base-pairs, respectively. The gene content and the relative positions of a total of 104 genes (70 peptide-encoding genes, 30 tRNA genes and four rRNA genes) are identical with the chloroplast DNA of the closely related species rice (*Oryza sativa*).

A detailed analysis of the two graminean plastomes allows the identification of hotspots of divergence which predominate in one region containing a cluster of tRNA genes and in two regions containing degenerated reading frames. One of these length differences is thought to reflect a gene transfer event from the plastome to the nucleus, which is followed by progressive degradation of the respective chloroplast gene resulting in gene fragments. The other divergent plastome region seems to be due to the complete loss of a plastid gene and its functional substitution by a nuclear encoded eukaryotic homologue.

The rate of neutral nucleotide substitutions is significantly reduced for protein coding genes located in the inverted repeat regions. This indicates that the existence of inverted repeat regions confers increased genetic stability of the genes positioned in these regions as compared to genes located in the two single copy regions.

Editing events cause the primary structures of several transcripts to deviate from the corresponding genomic sequences by C to U transitions. The unambiguous deduction of amino acid sequences from the nucleotide sequences of the corresponding genes is, therefore, not possible. A survey of the 25 editing positions identified in 13 different transcripts of the maize plastome shows that representatives of all protein coding gene classes are subject to editing. A strong bias exists for the second codon position and for certain codon transitions. Based on the number and the codon transition types, and taking into account the frequency of putative editing sites in all peptide encoding genes and unidentified reading frames, a total number of only few more than the experimentally verified 25 editing sites encoded in the maize plastome is estimated. This corresponds to 0.13% of amino acid positions which cannot be derived from the corresponding codons present in the corresponding genes.

Keywords: chloroplast; complete cpDNA sequence; *Zea mays*; genetic map; mRNA editing

**Corresponding author*

Introduction

Sequence analysis of complete chloroplast DNA (cp DNA) was first achieved in 1986 when the structures of the plastomes from tobacco (*Nicotiana tabacum*; Shinozaki *et al.*, 1986) and the liverwort *Marchantia polymorpha* (Ohyama *et al.*, 1986) were reported. Since then the complete sequences of the plastomes from rice (*Oryza sativa*; Hiratsuka *et al.*, 1989), *Euglena gracilis* (Hallick *et al.*, 1993), black pine (*Pinus thunbergii*; Wakasugi *et al.*, 1994) and from the

Abbreviations used: cp, chloroplast; IR, inverted repeat; SSC, LSC, small and large single copy; ORF, open reading frame; ACCase, acetyl-CoA-carboxylase.

0022-2836/95/350614-15 $12.00/0

root parasite *Epifagus virginiana* (Wolfe *et al.*, 1992) have been determined. Although sequence analyses of selected genes from the plastome of maize, as one of the major crop plants, were first published some 15 years ago (Schwarz & Kössel, 1979, 1980; McIntosh *et al.*, 1980; Koch *et al.*, 1981), and although the continued effort of several groups led to the structural and functional investigation of a large number of genes encoded by the maize plastome (for a survey see Rodermel & Bogorad, 1987, 1988, 1989), the complete sequence analysis of the maize plastome had not been accomplished.

The following considerations have led us to complete the nucleotide sequence determination of the maize plastome as a second chloroplast DNA from a monocotyledonous plant species.

In contrast to comparative analyses of complete plastome sequences from different plant lineages, a comparison of the complete cp DNA sequences of the closely related graminean species maize and rice was expected to allow the identification of evolutionary hotspots and to provide insights into specific mechanisms and traits underlying plastome divergence.

Editing events have been detected for several transcripts of the maize plastome, such as the transcripts of the genes *rpl*2 (Hoch *et al.*, 1991), *ndh*A (Maier *et al.*, 1992a), *ndh*B (Maier *et al.*, 1992b), *rpo*B (Zeltz *et al.*, 1993), *pet*B (Freyer *et al.*, 1993) and of the intron containing reading frame *ycf*3 (IRF170; Ruf *et al.*, 1994). As a consequence, single positions of certain mRNA sequences deviate from the corresponding genomic sequences by C to U transitions. Therefore, unambiguous deduction of amino acid sequences from the corresponding gene sequences, as had been tacitly implied for plastome sequences previously, is no longer possible (Neckermann *et al.*, 1994). The complete sequence analysis of the maize plastome, in combination with the experimental reconsideration of potential editing sites deduced from certain characteristics already evident from the limited number of previously identified editing sites was expected to allow an estimate of the total number of editing sites encoded in the transcripts of the maize plastome.

In this paper, we present the analysis of the complete chloroplast DNA sequence from maize and the genetic map derived from this sequence. Its comparison with the plastome sequence of rice reveals three areas of rapid divergence which apparently reflect different stages of gene deletions. Determination of neutral substitution rates provides evidence for the two large inverted repeat regions, which are characteristic of the majority of plastomes, acting as genetically stabilizing elements. This stabilizing effect is exerted specifically on the genes contained in the inverted repeat regions which is different from the stabilizing effect observed earlier on the anatomy of entire plastomes (Palmer & Thompson, 1982). A survey of the editing sites previously verified experimentally and of additional putative editing sites is presented. Experimental verification of some of the latter leads to a minimum estimate of the number of editing sites present in the maize plastome-encoded transcripts.

Results and Discussion

Size, structur and genes of the maize plastome

The chloroplast genome of *Zea mays* is a circular double-stranded DNA of 140,387 bp with an overall A + T content of 61.5%. As observed for other plastomes (Shimada & Sugiura, 1991), the A + T content is not distributed evenly. It tends to be higher in non-coding regions (average 71.2%), somewhat lower in the regions coding for peptides (60.5%) but considerably lower in regions coding for tRNAs (47.0%) and rRNAs (45.3%). The A + T content of the IR-region, amounts to only 56.0%, as a result of its coding for the rRNAs together with eight tRNAs, whereas the A + T contents of the LSC and SSC region are 63.7% and 67.2%, respectively.

The maize plastome (Figure 1A) harbours a pair of inverted repeat regions (IR_A and IR_B) consisting of 22,748 bp each, which is in good agreement with the size of 22 kbp estimated earlier from electron microscopic data (Bedbrook *et al.*, 1977). The inverted repeat regions are separated by a large single copy region (LSC) of 82,355 bp and a small single copy region (SSC) of 12,536 bp. The comparison of plastome structures and sizes (Figure 1B) indicates an overall genomic pattern which is similar in the majority of higher plant plastomes (Sugiura, 1992), although the four regions are subject to major size variations among various plant species (Palmer, 1991). The positions of all genes identified in the maize plastome are indicated in Figure 1A. This shows that the major portion of the maize plastome consists of coding regions (57.2%), whereas intergenic regions (including altogether 22 introns) comprise 42.8%. The percentage given for coding/intergenic regions, however, must be regarded as a minimal estimation since the presence of "hidden" translational start and stop codons created by RNA editing must be taken into consideration (Neckermann *et al.*, 1994). Moreover, the recent detection of a new structural RNA gene in the tobacco chloroplast genome (Vera & Sugiura, 1994) which, however, is not encoded in the maize plastome indicates that chloroplast genomes may contain additional hitherto unidentified genes. Of the 20 intron-containing genes (including the intron containing reading frame *ycf*3) five, *rpl*2, *ndh*B, *rps*12, *trn*I(GAU) and *trn*A(UGC), are located within the inverted repeat regions and, therefore, occur as duplicates. Consequently, the number of different intron-containing genes and introns is only 15 and 17, respectively. Of these introns only the one contained in the *trn*L(UAA) gene belongs to the group I introns, whereas all the remaining introns at the RNA level show the six domain secondary structure characteristic of group II introns (Michel & Dujon, 1983). The location of the first exon of the *rps*12 gene in the large

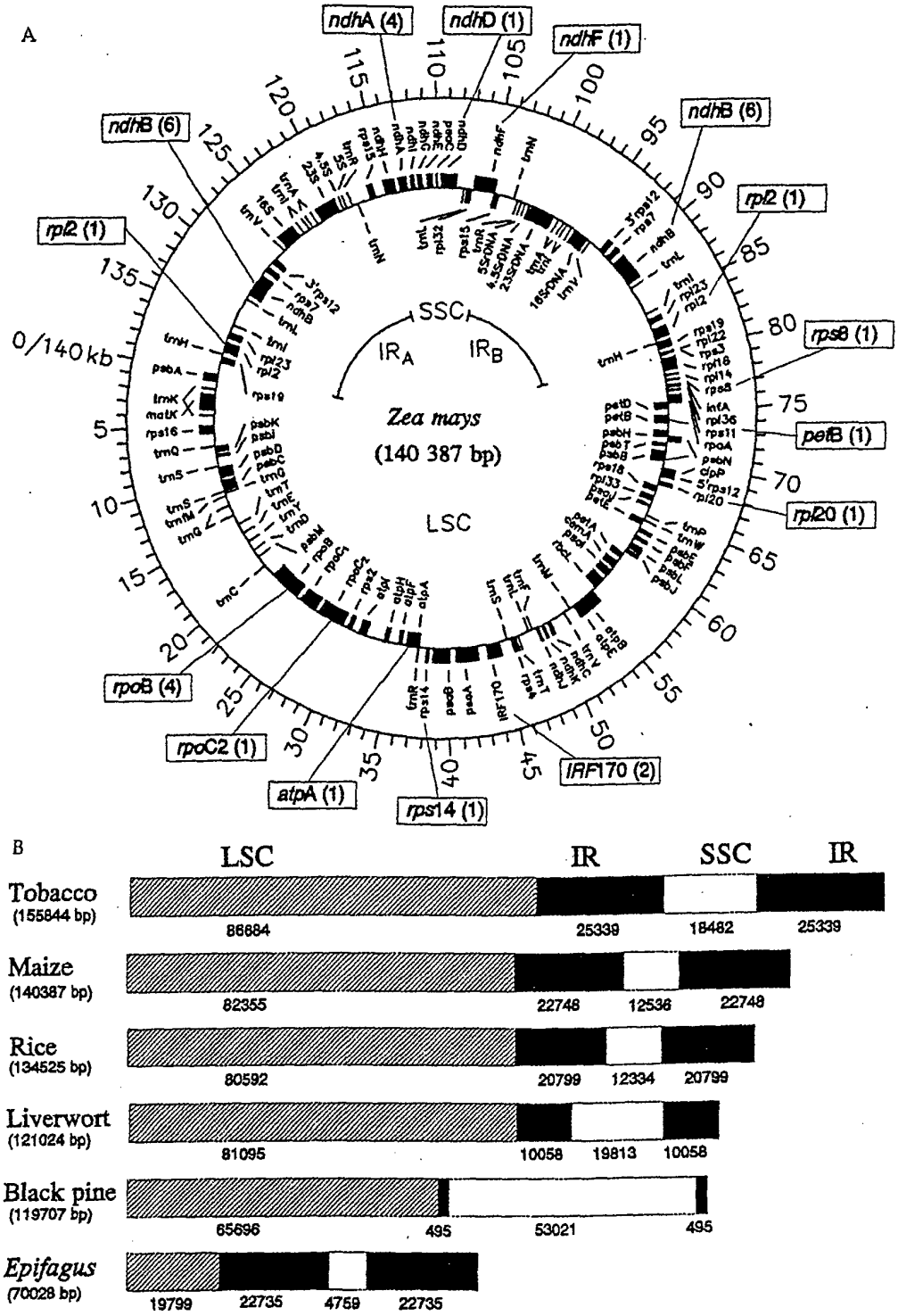


Figure 1. A, Gene organization of the *Zea mays* plastome. The inverted repeat regions IR_A and IR_B, respectively, divide the rest of the circular genome into large (LSC) and small (SSC) single copy regions. Genes drawn outside the circle are transcribed clockwise. Genes in which editing sites have been detected are marked by framing. The numbers within the parentheses behind the gene symbols indicate the numbers of editing sites observed in the respective genes. The orientation of the entire SSC region has been reversed as compared with the earlier version based on restriction site analysis (Larrinua *et al.*, 1983). This reorientation, representing one of the two possible structural isomers of a cp DNA molecule carrying IR-regions (Palmer, 1983), is now in accordance with the genetic map of the plastomes from tobacco, *Marchantia polymorpha*, rice and black pine, respectively. B, Length comparison of the completely sequenced higher plant plastomes. LSCs are marked by striations, SSCs by open boxes and IRs by filled boxes. Tobacco: Shinozaki *et al.* (1986); rice: Hiratsuka *et al.* (1989); *Marchantia polymorpha* (liverwort): Ohyama *et al.* (1986); black pine: Wakasugi *et al.* (1994); *Epifagus*: Wolfe *et al.* (1992).

Table 1. Genes contained in the maize plastome

16 S, 23 S, 4.5 S, 5 S ribosomal RNAs	*rrn*16, *rrn*23, *rrn*4.5, *rrn*5
tRNAs	*trn* (30 species, of which 6 contain an intron)
Proteins of the small ribosomal subunit	*rps*2, *rps*3, *rps*4, *rps*7, *rps*8, *rps*11, *rps*12*, *rps*14, *rps*15, *rps*16*, *rps*18, *rps*19
Proteins of the large ribosomal subunit	*rpl*2*, *rpl*14, *rpl*16*, *rpl*20, *rpl*22, *rpl*23, *rpl*32, *rpl*33, *rpl*36
Subunits (α, β, β′, β″) of the DNA-dependent RNA polymerase	*rpo*A, *rpo*B, *rpo*C1, *rpo*C2
Subunits of NADH-dehydrogenase	*ndh*A*, *ndh*B*, *ndh*C, *ndh*D, *ndh*E, *ndh*F, *ndh*G, *ndh*H, *ndh*I, *ndh*J, *ndh*K
Subunits of photosystem I and photosystem II	*psa*A, *psa*B, *psa*C, *psa*I, *psa*J *psb*A, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*I, *psb*J, *psb*K, *psb*L, *psb*M, *psb*N, *psb*T
Large subunit of rubisco	*rbc*L
Subunits of the cytochrome *b*/*f* complex	*pet*A, *pet*B*, *pet*D*, *pet*G
Subunits of ATP synthase	*atp*A, *atp*B, *atp*E, *atp*F*, *atp*H, *atp*I
Translational initiation factor I	*inf*A
Maturase	*mat*K
Protease	*clp*P
Envelope membrane protein	*cem*A

Intron-containing genes are marked by asterisks.

single copy region far away from its second and third exons, which are located as duplicates in the inverted repeat regions, requires a *trans*-splicing mechanism between exon I and exon II in order to produce mature *rps*12 transcripts (Fukuzawa *et al.*, 1986; Torazawa *et al.*, 1986; Ciese *et al.*, 1987).

As reported earlier for several gene clusters of the maize plastome and for the complete sequences of other plastomes (for reviews see Palmer, 1991; Sugiura, 1992), genes are encoded in both DNA strands (Figure 1A). It is also noteworthy that most of the genes are arranged within clusters of identical polarity, which often allows expression in the form of large polycistronic primary transcripts which are processed by splicing, cleavage to oligo- and monocistronic mRNAs and by editing (see below). In Table 1 the different gene classes and genes identified in the maize plastome are listed, whereas additional conserved unidentified open reading frames (ORFs) are summarized in Table 2. No homologues to the maize plastome ORFs could be found by data base searches (see Materials and Methods). Table 3 shows

Table 2. Comparison of conserved hypothetical chloroplast reading frames (*ycf*) encoded by the plastomes of maize, rice, tobacco, the liverwort *Marchantia polymorpha*, black pine, *Epifagus virginiana* and *Euglena gracilis*

	Maize	Rice	Tobacco	Liverwort	Black pine	*Epifagus*	*Euglena*
ycf1	—	—	ORF 1901	ORF 464/1068	ORF 1756	ORF 1738	—
ycf2	—	—	ORF 2280	ORF 2136	ORF 2054	ORF 2216	—
ycf3	IRF 170	IRF 170	IRF 168	IRF 168	IRF 169	—	—
ycf4	ORF 185	ORF 185	ORF 184	ORF 184	ORF 184	—	ORF 206
ycf5	ORF 321	ORF 321	ORF 313	ORF 320	ORF 320	—	—
ycf6	ORF 29	ORF 29	ORF 29	ORF 29	ORF 29	—	—
ycf7	ORF 31	ORF 31	ORF 31	ORF 31	ORF 62b	—	—
ycf8/*psb*T	ORF 33	ORF 35	ORF 34	ORF 35	ORF 35	—	ORF 31
ycf9	ORF 62	ORF 62	ORF 62	ORF 62	ORF 62	—	ORF 65
ycf10/*cem*A	ORF 230	ORF 230	ORF 229	ORF 434	ORF 261	—	—
ycf11/*acc*D	—	ORF 106	ORF 512	ORF 316	ORF 321	ORF 493	—
ycf12	—	—	—	ORF 33	—	—	ORF 33
ycf13	—	—	—	—	—	—	ORF 485
ycf14/*mat*K	ORF 544	ORF 542	ORF 509	ORF 370	ORF 515	ORF 439	—
ycf15	—	—	ORF 70	—	—	—	—
	ORF 148	ORF 70					
	ORF 75	ORF 82					
	ORF 69	ORF 91					
	ORF 137	ORF 137					
	ORF 85	ORF 85					
	ORF 23	ORF 23					
	ORF 133	ORF 133					
	ORF 49	ORF 109					
	ORF 63	ORF 63					

In the lower part, ORFs conserved only between the plastomes of maize and rice are given. Positions of the reading frames can be retrieved from the supplement of the EMBL data entry (accession no. X86563).

Table 3. Codon usage of the maize plastome

TTT	Phe	gaa	699	65%	TCT	Ser	gga	370	28%	TAT	Tyr	gua	538	79%	TGT	Cys	gca	152	73%
TTC	Phe	gaa	383	35%	TCC	Ser	gga	264	20%	TAC	Tyr	gua	139	21%	TGC	Cys	gca	57	27%
TTA	Leu	uaa	691	34%	TCA	Ser	uga	221	16%	TAA	End	*	38	46%	TCA	End	*	24	29%
TTG	Leu	caa	372	18%	TCG	Ser	uga	123	9%	TAG	End	*	21	25%	TGG	Trp	cca	336	100%
CTT	Leu	uag	434	21%	*CCT*	Pro	ugg	310	38%	CAT	His	gug	315	72%	CGT	Arg	acg	263	23%
CTC	Leu	uag	130	6%	*CCC*	Pro	ugg	192	24%	CAC	His	gug	120	28%	*CGC*	Arg	acg	99	9%
CTA	Leu	uag	289	14%	CCA	Pro	ugg	212	26%	CAA	Gln	uug	496	76%	*CGA*	Arg	acg	244	22%
CTG	Leu	uag	115	6%	CCG	Pro	ugg	99	12%	CAG	Gln	uug	154	24%	*CGG*	Arg	acg	83	7%
ATT	Ile	gau	766	50%	ACT	Thr	ggu	409	41%	AAT	Asn	guu	552	75%	AGT	Ser	gcu	271	20%
ATC	Ile	gau	278	18%	ACC	Thr	ggu	189	19%	AAC	Asn	guu	184	25%	AGC	Ser	gcu	96	7%
ATA	Ile	cau	476	31%	ACA	Thr	ugu	279	28%	AAA	Lys	uuu	682	73%	AGA	Arg	ucu	325	29%
ATG	Met	cau	430	100%	ACG	Thr	ugu	119	12%	AAG	Lys	uuu	252	27%	AGG	Arg	ucu	110	10%
GTT	Val	gac	423	38%	*GCT*	Ala	ugc	522	45%	GAT	Asp	guc	526	78%	GGT	Gly	gcc	442	32%
GTC	Val	gac	133	12%	*GCC*	Ala	ugc	174	15%	GAC	Asp	guc	147	22%	GGC	Gly	gcc	150	11%
GTA	Val	uac	398	36%	GCA	Ala	ugc	335	29%	GAA	Glu	uuc	752	76%	GGA	Gly	ucc	542	39%
GTG	Val	uac	147	13%	GCG	Ala	ugc	131	11%	GAG	Glu	uuc	242	24%	GGG	Gly	ucc	257	18%

The codon usage has been deduced from all protein-coding genes and the ORFs summarized in Table 2. Codons are given in capital, anticodons in lower case letters. Stop codons are marked by an asterisk (*). Codons said to be translated by unconventional base-pairing are indicated in italics. Beside the total numbers of the individual codons the percentages of individual codons with respect to all codons of a family specifying the same amino acid, are also given.

the codon usage of the mRNAs encoded in the maize plastome and the anticodons present in the 30 tRNA species. The high A + T content is reflected in the codon usage with a strong bias for codons containing A or U residues in the third positions, in accordance with the codon usage of other plastomes (Shimada & Sugiura, 1991). It has been suggested that the 30 tRNA species also encoded in the plastomes of several other higher plants are sufficient to read all 61 amino acid codons (Maréchal-Drouard *et al.*, 1991) and that, therefore, no import of nuclear-encoded tRNAs is necessary to complement this set of 30 tRNA species (Sugiura, 1992).

Divergence between the maize and rice plastome sequences

Both maize and rice belong to the graminean family of monocotyledonous plants. The availability of the complete plastome sequences from both species opens the possibility of identifying hotspots of divergence within a background of otherwise highly conserved sequences of two closely related plastomes.

The size difference of 5863 bp between the two plastomes is caused mainly by a limited number of highly divergent regions. One of these is within the inverted repeats (Figure 2A), containing the large open reading frame *ycf2* (ORF2280/2136/2054/2216 in the tobacco, liverwort, black pine and *Epifagus* plastome, respectively). It is reduced by various deletions to a series of shorter reading frames in the plastomes of maize and rice. Since *ycf2* is encoded in the plastome of the non-photosynthetic parasitic flowering plant *Epifagus virginiana* (see Table 2), which lacks all genes of photosynthetic metabolism and chlororespiration, *ycf2*-encoded proteins seem not to be involved in these processes. This is supported by the detection of a higher level of *ycf2*-encoded protein in non-photosynthetic tissues of tomato plants (Richards *et al.*, 1994). On the basis of limited sequence similarities to ATPases of the CDC48 family it seems possible that Ycf2 protein is a plastid-specific ATPase (Wolfe, 1994).

The elimination of *ycf2*, which accounts for a size difference of 2181 bp for each of the inverted repeat regions, is more advanced in the rice plastome as compared to the maize plastome. The fragmentation of *ycf2* probably reflects different stages of gene deletion after its original function has been taken over by a *ycf2* copy transferred to the nuclear genome. It is generally assumed that a large number of similar gene transfer events is a major reason for the size reduction of the genome of the original bacterial endosymbiont to the genomes of the present day chloroplasts (Palmer, 1991; Gray, 1991). Thus, the remnants of *ycf2* still recognizable in the maize and rice plastomes probably provide a representative example of the transition stages of chloroplast gene deletion, which is believed to have occurred frequently during the evolution of chloroplast genomes. Since *ycf2* could be detected in chloroplast DNA of two other monocot species that do not belong to the graminean family, *Lilium elegans* and *Allium cepa* (Katayama & Ogihara, 1993), loss of functional plastome-encoded *ycf2* must have occurred in the ancestor of the grasses. The subsequent successive fragmentation of the reading frame becomes obvious by comparing the corresponding plastome sequences of the two grasses, maize and rice.

A second hotspot of divergence is located in the large single copy region between the genes *rbc*L and *cem*A. As depicted in Figure 2B, the *acc*D gene that encodes one of the subunits of the prokaryotic form of acetyl-CoA-carboxylase (ACCase) present in the plastome of tobacco (ORF512), black pine (ORF321) and liverwort (ORF316) is reduced to a reading frame of only 106 codons (ORF106) in the rice plastome and has shrunk to zero in the maize plastome. The complete deletion in maize again strongly suggests that ORF106 in rice represents an intermediate stage in the degenerating *acc*D gene. However, while the fragmentation of *ycf2* is more progressed in the rice

plastome, the deletion of *accD* and its flanking regions is only partially advanced in the rice plastome but is complete in the maize plastome.

In contrast to the deletion of the plastome-encoded *ycf2*, the function of which has been probably taken over by a copy transferred to the nuclear genome, *accD* seems to have been completely lost in maize. Dicotyledonous plants have been shown to contain two forms of ACCase: a prokaryotic form consisting of three protein components and a eukaryotic form consisting of three functional domains within a single protein (Konishi & Sasaki, 1994). The localization of the prokaryotic form in pea chloroplasts could be shown by identifying one of the subunits, the plastome-encoded AccD protein (Sasaki *et al.*, 1993), whereas the eukaryotic form was detected outside the plastids (Konishi & Sasaki, 1994). Herbicides of the diphenoxypropionic acid type and the cyclohexadione type (graminicides) selectively inhibit the plastid-localized *de novo* fatty acid synthesis in some graminean species (including maize, rice and wheat) while other monocotyledonous families and dicotyledonous plant species are herbicide-resistant. Whereas the eukaryotic ACCase from pea is inhibited by graminicides, the prokaryotic ACCase is herbicide- resistant (Konishi & Sasaki, 1994). From this it can be concluded that the herbicide-sensitive graminean species have lost

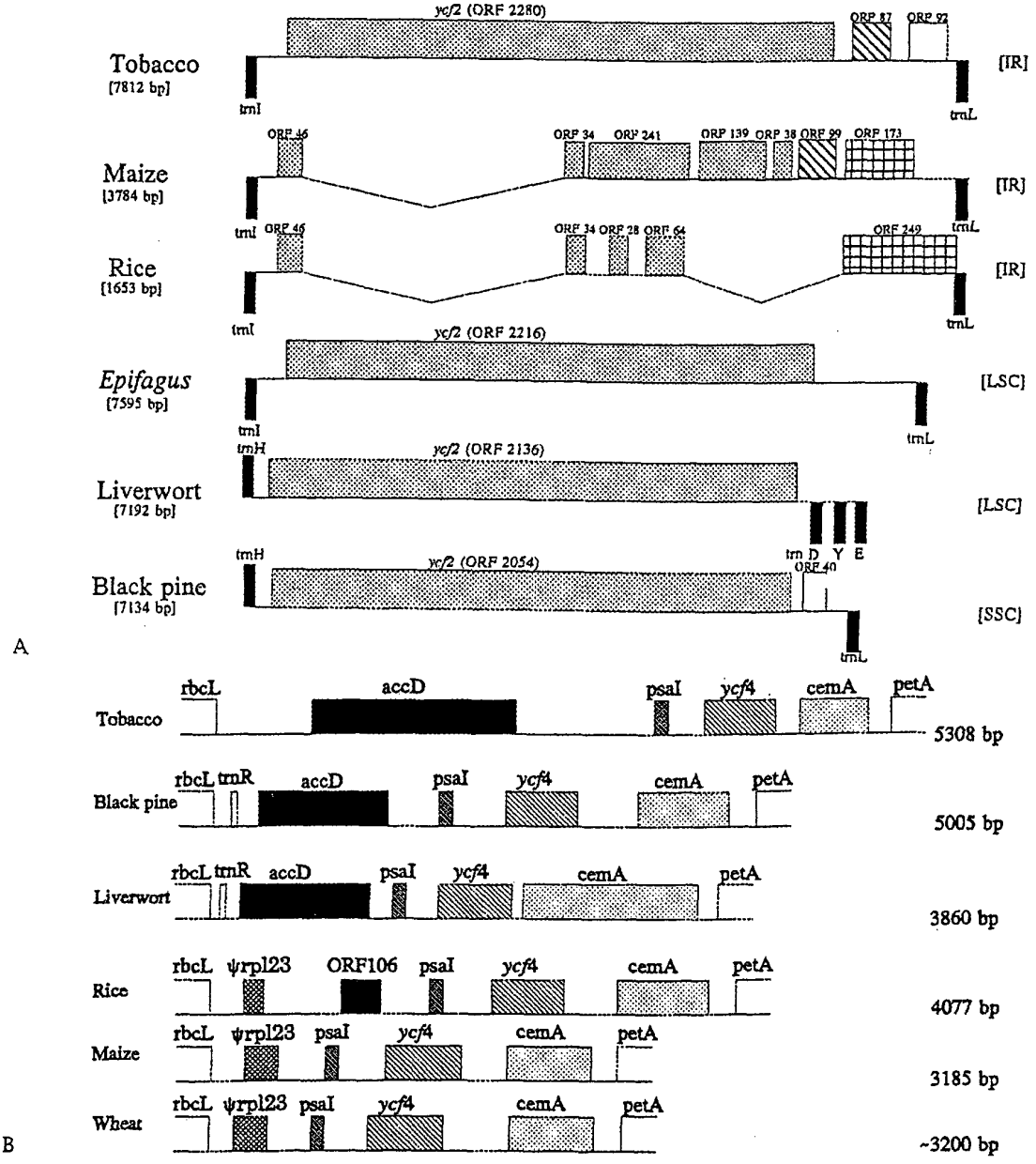


Figure 2A and B

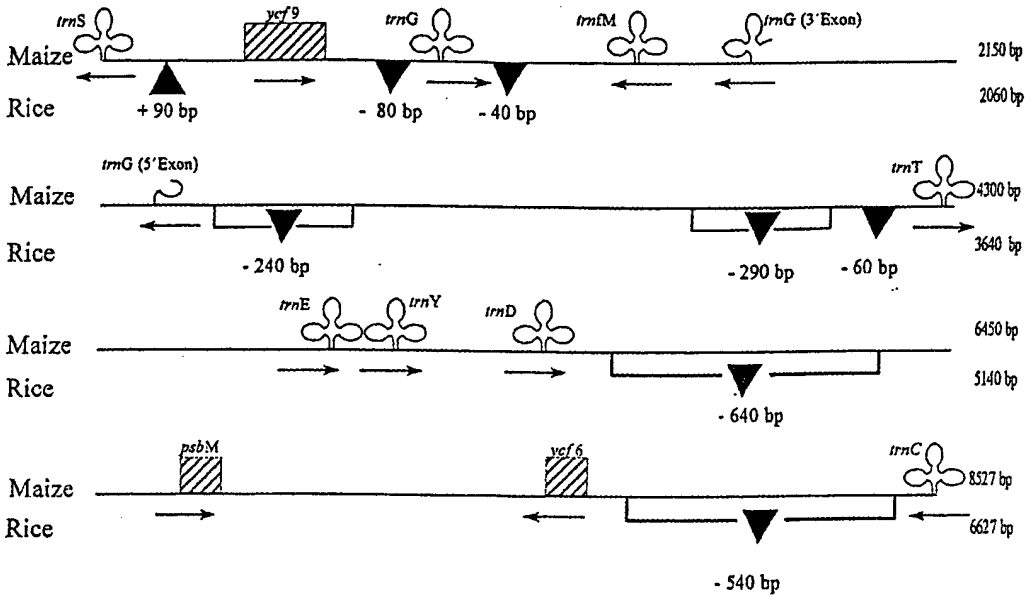


Figure 2C

Figure 2. A, Evolutionary loss of *ycf2* within the graminean plants maize and rice. Homologous regions are marked by the same pattern. Open reading frames and genes above the baseline are transcribed left to right. B, Disappearance of the *acc*D gene within the graminean plants rice, maize and wheat (Ogihara *et al.*, 1991). The *acc*D gene from tobacco (marked by a filled box) is reduced from 512 amino acids (aa) to 321 aa and 316 aa in black pine and the liverwort *Marchantia polymorpha*, respectively. Remnants of only 106 aa are detectable in rice, whereas in maize and wheat a homologous open reading frame is missing. C, Comparison of a tRNA-rich region between maize and rice within the LSC. Upward-pointing triangles indicate insertions in the rice sequence compared to maize. Deletions in the rice sequence are marked by downward-pointing triangles. Arrows indicate the direction of trancription of the respective gene.

the prokaryotic form of the enzyme. ACCase activity in their plastids must have been taken over by an enzyme of the eukaryotic type. This requires a eukaryotic type ACCase, carrying a plastid targeting signal sequence, encoded in the nuclear genome of herbicide-sensitive Gramineae. Interestingly, in maize two ACCase isoforms of the eukaryotic type could be purified, from which only one was found to be localized in chloroplasts (Egli *et al.*, 1993). The cause for the deletion of plastome-encoded *acc*D in maize and related graminean species, therefore, is likely not to be the transfer to the nuclear genome and subsequent activation of the transferred organellar gene copy in the nucleus but the substitution of the organellar gene function by a nuclear-encoded gene of eukaryotic origin. Recently, Bubunenko *et al.* (1994) could show that in spinach, where the plastome-encoded *rpl*23 reading frame is disrupted, the organellar L23 protein has been functionally replaced by its cytosolic nuclear-encoded ribosomal L23 homologue.

The plastome region between *rbc*L and *cem*A shows length polymorphisms even between the two closely related monocotyledonous species *Aegilops crassa* and *Aegilops sqarossa* (Ogihara *et al.*, 1991). To investigate if sequence differences in this very fast evolving plastome region may even exist at the subspecies level, we determined the corresponding sequence from a teosinte species. No sequence differences could be detected between the two subspecies maize (*Zea mays* subsp. *mays*) and teosinte (*Zea mays* subsp. *mexicana*), supporting the hypothesis that a teosinte-like species was one of the more recent relatives of maize (Doebley & Stec, 1991). *Zea mays*, *Zea diploperennis* and *Zea luxurians* also show no sequence deviation within this variable plastome region (Morton & Clegg, 1993). This is consistent with the hypothesis that *Zea* is a relatively young genus (Doebley *et al.*, 1987).

The pseudogene ψ*rpl*23, located between *rbc*L and ORF106 in the rice plastome, is maintained in the homologous position of the maize plastome (Figure 2B) in a less degenerated form, by which the size reduction caused by the loss of ORF106 is compensated to some extent. ψ*rpl*23, present in the divergent region, is not indicative of a gene transfer to the nuclear genome or the complete loss of a plastome-encoded gene as suggested above for *ycf*2 and *acc*D, since functional copies of the *rpl*23 gene still exist in the two inverted repeat regions. Altogether this region contributes to a 844 bp size reduction of the maize plastome as compared to the rice plastome.

A third hotspot of divergence can be recognized in the region of the large single copy region containing the tRNA gene cluster *trn*S(UGA), *trn*G(GCC), *trn*fM(CAU), *trn*G(UCC), *trn*T(GGU), *trn*E(UUC), *trn*Y(GUA), *trn*D(GUC), *trn*C(GCA). In this region, no differences in gene content and gene order can be detected between the maize and rice plastomes

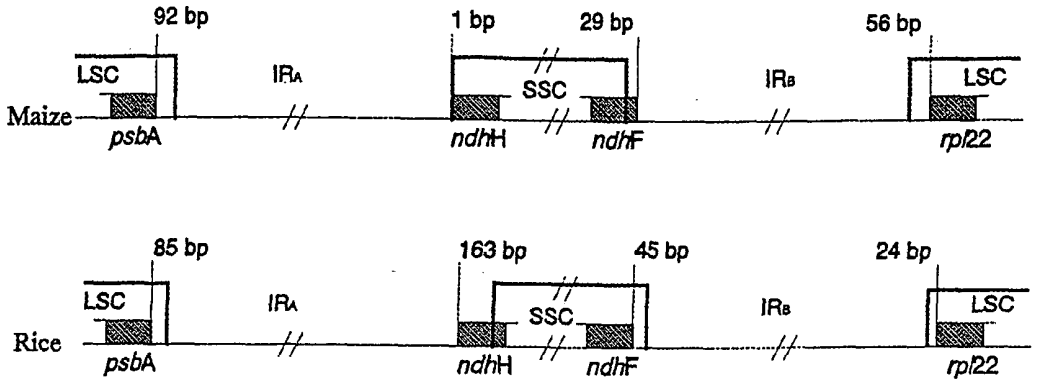


Figure 3. Comparison of the relative positions of junctions LSC/IR_A, IR_A/SSC, SSC/IR_B, IR_B/LSC and *psbA*, *ndhH*, *ndhF* and *rpl22*, respectively, between the plastomes of maize and rice. The number of base-pairs between the given genes and the respective borders are indicated.

(Figure 2C). However, as a result of a large number of deletion/insertion events of variable lengths within several of the intergenic regions, sequence and size divergence is markedly enhanced in the entire area. A size difference between maize and rice plastomes of altogether 1900 bp is caused by these events. No apparent reasons for this region being subject to so many alterations can be recognized. It is tempting to speculate that the clustering of tRNA genes *per se* may somehow be causative for structural instability. On the other hand, the mere prevalence of intergenic sequences within the cluster of the tRNA genes (87.7% of this region consists of intergenic sequences as opposed to an average of only 41.8% in the entire plastome) may also contribute to the higher rate of deletion/insertion events.

The borders between the two inverted repeat regions (IR_A and IR_B) and the two single copy regions (LSC and SSC) usually differ among various plastome species. Accordingly, large expansions (and reductions) of plastome sizes are often caused by expansions (or reductions) of the inverted repeat regions (for a review, see Palmer, 1991). Detailed analyses of the border positions have been presented (Sugita *et al.*, 1984; Moon & Wu, 1988; Prombona & Subramanian, 1989; Maier *et al.*, 1990). In Figure 3 the exact IR-border positions, with respect to the adjacent genes from the maize and rice plastomes, are compared. This comparison demonstrates that all the border positions can vary even between plastomes of closely related species such as maize and rice. Whereas very small shifts of only 7 and 32 bp are observed for the two borders with the large single copy region, somewhat larger shifts of 162 and 74 bp have occurred at the borders of the IR with the small single copy region. It should also be noted that two of the borders are located within coding regions. In maize the initiation codon of the *ndh*H gene is disrupted by the IR_A/SSC border (Maier *et al.*, 1990) and the SSC/IR_B border is located within the *ndh*F coding region. In rice the IR_A/SSC border is shifted into the central region of the *ndh*H gene, whereas the entire *ndh*F gene has moved into the small single copy region. The shift of the IR_B/SSC and SSC/IR_A borders in rice compared to maize has probably been mediated during evolution by a mechanism of

A

```
                                        IRB   SSC (Maize)
        Stop ndhF                         ↓
Maize TTCATGGAATAAAGAAT-AAG-AAC-AAGGATA----------------------          ------------ 30
Rice  ............T....T...G..GA.......TGAGTCTACCGCTTACAAGAA   126 nt   GAATGGAAAAAA 198


Maize ------------------------------------------------------------------------CA  32
Rice  TCGGCAAATAGGAAAGCTACTTAGGCAGGAGATAGGGGAATTCCTTAAGAAAGAAAAAAGAATAAGAACACAGATA.. 270
          ↑                                          Stop ndhF
    IRB    SSC  (Rice)
```

B

```
Maize ATTTTTTGCAATT----------------------          ------------------------GTAAGGGG  21
Rice  ........A...GAATTGAATGAAGAAGAATTGG   31 nt   AGGTCCCTTCGAAATTGCATTATT........  97
```

Figure 4. In A, comparison of the nucleotide sequences of the region containing the 3'-terminal part of *ndh*F and the IR_B/SSC junction from the plastomes from maize and rice is shown. Nucleotides of the non-coding strand of *ndh*F are given in italics. The first nucleotide of the IR_B region is marked by an arrow. The regions showing interrupted direct repeat sequences are underlined. In B the nucleotide sequences of parts of the intergenic region between *ycf4* (ORF185) and *cem*A are compared. The regions of the two octanucleotide direct repeats are marked by underlining. Positions identical with the maize sequence are indicated by dots, deleted regions are marked by dashes.

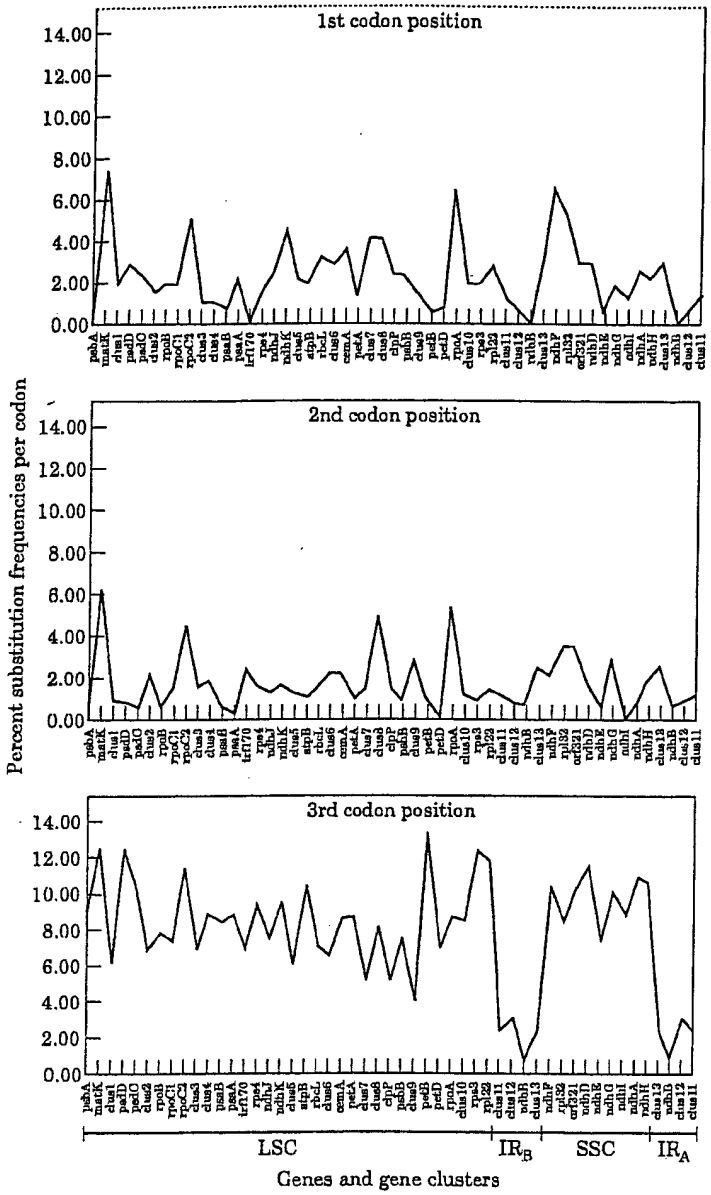


Figure 5. Frequency of nucleotide substitutions in first, second and third codon positions between homologous protein-coding genes and conserved ORFs from the maize and rice plastome. Protein-coding genes and in the case of coding regions smaller than 100 codons, clusters of protein coding genes (clus) were plotted against the substitution frequency.

intramolecular recombination between two short direct repeat sequences in the plastome of a maize ancestor, which would have contained an arrangement of the IR/SSC borders similar to the rice plastome (Figure 4A; Maier *et al.*, 1990). Intramolecular recombination between short direct repeat sequences could be proposed as one cause of length differences between the two closely related plastomes of maize and rice. As shown in Figure 4B, recombination of two octanucleotide direct repeats in the intergenic region between *ycf4* (ORF185) and *cem*A of the "rice-like" plastome of a progenitor of maize is probably the cause of an 81 base-pair deletion in maize as compared to rice. Slippage of DNA polymerase during plastome replication seems to be an additional mechanism causing length mutations (Palmer, 1991; Morton & Clegg, 1993).

The inverted repeat regions confer increased sequence stability

Comparative analyses of conserved restriction sites and sequences of selected genes among several plastomes have led to the suggestion that sequences contained in the inverted repeat regions diverge at a slower rate as compared to sequences located in the single copy regions (Palmer, 1991; Wolfe *et al.*, 1987). As an independent test of this proposal we have analysed the frequency of neutral nucleotide substitutions between all of the homologous protein genes encoded in the plastomes of maize and rice. In Figure 5 the substitution rates of individual genes, in the case of small genes the substitution rates of gene groups, are plotted according to their map positions in the two plastomes. Not unexpectedly the first and

second codon positions show a significantly lower frequency of around two substitutions per some hundred codons, whereas the third codon positions are substituted at a three- to fivefold higher rate. The latter, however, drops to almost the same low level as the first and second codon positions for the genes positioned in the inverted repeat regions. Thus, a significant reduction of the neutral substitution rate by a factor of 2 to 3 becomes evident for the genes located within the inverted repeat regions. This clearly shows that a stabilizing effect is conferred to the sequences in the inverted repeat regions and thus confirms the conclusion drawn earlier.

Fine tuning of the genetic information by transcript editing

C to U editing events occurring in the transcripts of several maize chloroplast genes have been observed (Hoch *et al.*, 1991; Maier *et al.* 1992a,b; Zeltz *et al.*, 1993; Freyer *et al.*, 1993; Ruf *et al.*, 1994). The positions of these genes and the numbers of edited positions identified in the respective transcripts are indicated in Figure 1A. A number of editing sites could also be detected at homologous and non-homologous positions of transcripts encoded in the plastomes of other plant species such as barley (Zeltz *et al.*, 1993), tobacco and spinach (Kudla *et al.*, 1992; Bock *et al.*, 1993; Neckermann *et al.*, 1994; Hirose *et al.*, 1994), snapdragon (Neckermann *et al.*, 1994), bell pepper (Kuntz *et al.*, 1992) and black pine (M. Sugiura, personal communication). Thus, editing has to be regarded as a third common step of RNA processing in plastids in addition to splicing and cleavage of polycistronic to oligo- and monocistronic transcripts.

In Table 4 the codon transitions caused by editing in the chloroplast transcripts of maize are shown. From the number of 25 different editing sites in maize (18 previously identified and seven additional ones identified in this work), a strong bias for the second codon position and for certain codon transitions is clearly detectable. All the editing events are C to U transition and, contrary to editing in plant mitochondrial transcripts where also reverse U to C editing is occasionally observed, no U to C transition could be detected for chloroplast transcripts so far. Of the 18 different sites detected in maize chloroplast transcripts previously only one (CAU to UAU) changes the first codon position and no editing of third codon positions could be identified. The most frequently observed transitions (UCA to UUA and CCA to CUA) convert serine and proline codons to leucine codons, whereas no alanine to valine or threonine to isoleucine transition, which would also be possible by C to U editing of second codon positions, have been detected so far. As initially realized for the "missing" initiation codon of the maize *rpl*2 gene (Hoch *et al.*, 1991), editing events occurring at internal codons of chloroplast mRNAs restore codons for amino acid residues, which are conserved at the DNA level in other species (Maier *et al.*, 1992a). For this reason editing can also be regarded as a genetic repair process acting at the transcript level. An alignment of all known chloroplast editing sites with homologous sequences of other plastomes shows that the peptide-encoding genes of the *Marchantia polymorpha* plastome already encode the respective amino acid residues at the gene level. Therefore, editing appears to be unnecessary for the transcripts encoded in the *Marchantia* plastome. This has led to the suggestion that editing is probably non-existent in the genetic system of *Marchantia* chloroplasts (a situation which is paralleled by the apparent absence of editing in the mitochondria of *Marchantia*; Oda *et al.*, 1992). The amino acid sequences encoded in the *Marchantia* plastome can, therefore, serve as a reference system for the screening of putative editing sites, and thus for providing a rough estimate of the number of editing sites encoded in a known plastome sequence.

The result of such a screening for potential C to U editing sites in all peptide-encoding genes of the maize plastome is presented in Table 5. In the upper portion of this Table the genes whose transcripts have already been analysed experimentally by cDNA sequencing, including the positions of the 18 editing sites identified previously, are listed. On the other hand, a large number of potential editing sites exist where no difference between the genomic and cDNA sequences was detected. The transcripts of several genes (e.g. *rpo*A, *pet*D, *rps*4, *rpo*C_1) in spite of putative editing sites, seem not to be subject to editing at all. A strong preference for certain codon transitions among the verified editing sites exists,

Table 4. Codon transitions caused by editing of maize chloroplast transcripts

First base of codon	Second base of codon U		C		A	G	Third base of codon
U	Phe		Ser	→	Tyr	Cys	U
	Phe	←1x	Ser		Tyr	Cys	C
	Leu	←12x	Ser	1x	Stop	Stop	A
	Leu	←4x	Ser		Stop	Trp	C
C	Leu		Pro	←	His	Arg	U
	Leu		Pro		His	Arg	C
	Leu	←4x	Pro		Gln	Arg	A
	Leu	←1x	Pro		Gln	Arg	C
A	Ile		Thr		Asn	Ser	U
	Ile		Thr		Asn	Ser	C
	Ile		Thr		Lys	Arg	A
	Met	Start 1x / Int. ←1x	Thr		Lys	Arg	G
G	Val		Ala		Asp	Gly	U
	Val		Ala		Asp	Gly	C
	Val		Ala		Glu	Gly	A
	Val		Ala		Glu	Gly	G

The arrows show the direction of the codon transitions with the numbers above the arrows indicating the observed frequencies of the respective transitions. One threonine (Thr) to methionine (Met) transition creates an initiator codon (Start), one (Int.) affects the formation of an internal methionine codon.

Table 5. List of all potential (see text) and experimentally verified editing sites in protein coding sequences from the plastome of maize

rpl2(1)	215	1	T->M	AoG	+
		104	H->Y	oAT	-
		123	S->L	ToA	-
		153	A->V	GoT	-
		170	A->V	GoC	-
ndhA(2)	362	17	S->L	ToG	+
		26	L->F	oTT	-
		31	P->S	oCC	-
		79	T->I	AoT	-
		82	L->F	oTT	-
		104	A->V	GoA	-
		126	L->F	oTT	-
		158	S->L	ToA	+
		188	S->L	ToA	+
		357	S->F	ToC	+
		361	S->L	ToA	-
rpoA(2)	205	37	L->F	oTT	-
ndhB(3)	510	25	L->F	oTT	-
		26	L->F	oTC	-
		28	H->Y	oAT	-
		52	S->F	ToT	-
		103	L->F	oTT	-
		156	P->L	CoA	+
		196	H->Y	oAT	+
		204	S->L	ToA	+
		233	L->F	oTT	-
		246	P->L	CoA	+
		247	A->V	GoC	-
		277	S->L	ToA	+
		298	L->F	oTT	-
		494	P->L	CoA	+
petB(4)	215	76	A->V	GoC	-
		129	A->V	GoA	-
		163	A->V	GoT	-
		204	P->L	CoA	+
petD(4)	160	138	A->V	GoA	-
rpoB(5)	1075	11	T->I	AoA	-
		40	T->I	AoA	-
		63	S->L	ToG	-
		96	T->I	AoC	-
		98	S->F	ToT	-
		156	S->L	ToG	+
		182	S->L	ToA	+
		183	S->L	ToA	-
		187	S->L	ToG	+
		206	P->L	CoG	+
		313	R->C	oGT	-
		358	P->S	CoA	-
		439	T->I	AoC	-
		564	S->L	ToT	-
		724	H->Y	oAT	-
		873	S->L	ToG	-
		1005	H->Y	oAT	-
ycf3(6)	170	13	S->F	ToC	+
		62	T->M	AoG	+
rps4(6)	201	26	P->L	CoT	-
		155	S->F	ToC	-
rpoC1(7)	683	343	T->M	AoG	-
		356	L->F	oTT	-
		469	T->I	AoT	-
		604	P->S	CoC	-
rpl23(7)	93	68	H->Y	oAT	-
		81	P->S	oCG	-
rps18	170	58	P->S	CoG	nd
		119	Q->*	oAA	-
atpF	183	6	H->Y	oAT	-
		7	S->F	ToT	-
		168	T->M	AoG	-
rpoC2	1527	5	A->V	GoC	-
		10	H->Y	oAT	-
		309	R->W	oGG	-
		311	P->L	CoC	-
		353	T->I	AoT	-
		444	H->Y	oAT	-
		463	P->S	oCC	-
		474	P->L	CoA	-
		495	S->L	ToG	-
		497	H->Y	oAC	-
		561	S->F	ToC	-
		822	H->Y	oAT	-
		886	H->Y	oAT	-
		912	R->W	oGG	-
		925	S->L	ToG	+
		965	H->Y	oAT	-
		971	T->I	AoT	-
		1000	A->V	GoT	-
		1002	L->F	oTT	-
		1040	P->S	oCC	-
		1070	T->I	AoC	-
		1119	L->F	oTC	-
		1131	S->F	ToT	-
		1185	P->L	CoA	-
		1193	L->F	oTC	-
		1274	S->L	ToG	-
		1320	P->L	CoC	-
ndhH	393	165	A->V	GoC	nd
		301	P->S	oCC	-
		356	P->S	oCA	-
ycf4	185	12	L->F	oTT	-
		78	S->F	ToC	-
		136	L->F	oTT	-
		138	P->S	oCT	-
		164	P->L	CoA	-
		181	P->S	oCA	-
ycf10 (cemA)	230	26	S->F	ToT	-
		41	T->I	AoC	-
		46	T->I	AoT	-
		79	T->I	AoA	-
		111	S->L	ToA	-
		181	L->F	oTT	-
		184	A->V	GoT	-
rpl12	124	119	A->V	GoC	-
		122	P->S	oCA	-
rps7	156	6	T->I	AoT	-
		113	P->S	oCG	-
		152	L->F	oTT	-
psaA	320	17	S->L	ToG	-
		121	P->S	oCC	-
		136	S->F	ToT	-
		255	S->L	ToG	-
ndhD	500	19	S->L	ToA	-
		85	L->F	oTT	nd
		293	S->L	ToA	+
		450	P->L	CoA	-
ndhI	180	3	P->S	oCT	-
		47	S->L	ToA	-
		161	S->L	ToG	-
		178	T->I	AoA	-
ndhG	176	6	P->S	oCA	-
		8	H->Y	oAT	-
		41	L->F	oTT	-
		110	S->L	ToA	-
		124	L->F	oTT	-
ndhF	738	8	A->V	GoC	-
		14	L->F	oTC	-
		16	L->F	oTT	-
		17	P->L	CoA	-
		21	S->L	ToA	+
		28	L->F	oTT	-
		33	T->I	AoA	-
		36	L->F	oTT	-
		47	L->F	oTC	-
		59	S->F	ToT	-
		117	H->Y	oAC	-
		225	L->F	oTT	nd
		281	L->F	oTT	nd
		583	L->F	oTT	nd
		411	L->F	oTT	nd
		436	S->F	ToT	nd
		537	S->F	ToC	-
		583	P->L	CoT	-
		697	T->I	AoC	nd
rps14	103	27	S->L	ToA	+
		98	A->V	GoA	-
rpl20	119	103	S->L	ToA	+
		108	L->F	oTT	-
atpI	248	11	L->F	oTC	-
		53	S->L	ToA	-
		146	A->V	GoG	-
atpA	507	2	A->V	GoA	nd
		127	S->F	ToT	nd
		383	S->L	ToA	+
		430	P->S	oCA	-
		464	H->Y	oAC	-
		503	L->F	oTT	-
ndhE	101	85	H->Y	oAT	-
		96	S->F	ToC	-
rps8	136	61	S->L	ToA	+
		108	A->V	GoT	nd
rps3	227	17	T->I	AoA	nd
		88	H->Y	oAT	-
		185	A->V	GoC	nd
ycf14 (matK)	544	208	S->F	ToT	nd
		336	L->F	oTT	nd
		387	S->L	ToT	-
		403	A->V	GoA	-
		419	S->L	ToG	-
		437	L->F	oTT	nd
		467	L->F	oTT	nd
		469	L->F	oTT	nd
ycf5	321	100	T->I	AoA	-
		115	S->L	ToG	-
psbK	61	16	L->F	oTT	-
		18	P->L	CoG	-
		46	L->F	oTC	-
ndhK	248	28	P->S	oCG	nd
		40	S->L	ToA	-
		46	L->F	oTT	-
		197	S->F	ToT	nd
		204	L->F	oTT	nd
		220	L->F	oTC	nd
psbN	43	10	S->F	ToC	-
ndhC	120	19	L->F	oTT	nd
		21	P->S	oCT	nd
		41	L->F	oTT	nd
rpl22	148	72	L->F	oTC	nd
rpl33	66	60	T->I	AoT	nd
psbH	73	5	T->I	AoC	nd
psbD	353	136	P->S	oCT	nd
psbC	473	177	L->F	oTC	nd
		246	L->F	oTT	nd
psbB	508	99	A->V	GoG	nd
ycf7	31	5	T->I	AoT	nd
		19	P->L	CoA	nd
		20	A->V	GoT	nd
psaA	751	13	A->V	GoT	nd
		320	L->F	oTT	nd
		713	T->I	AoT	nd
psaB	735	226	L->F	oTT	nd
		258	L->F	oTT	nd
		437	P->L	CoT	nd
atpE	137	39	P->S	oCC	nd
		97	L->F	oTT	nd
		119	A->V	GoT	nd
		120	L->F	oTC	nd
		135	P->S	oCA	nd
rps11	143	90	T->I	AoA	nd
atpB	498	36	P->S	oCC	nd
		378	R->W	oGG	nd
rpl32	59	49	P->S	oCA	nd
rbcL	476	418	A->V	GoA	nd
clpP	216	48	T->I	AoA	nd
rps2	236	42	T->I	AoT	nd
		43	T->I	AoA	nd
ycf9	62	60	L->F	oTT	nd

The numbers of encoded amino acids are given for each protein-coding gene together with the codon positions where conservation with the respective amino acids encoded by the liverwort plastome can be achieved by C to U transitions (lower case). Experimentally verified editing positions are marked by +, position where no C to U editing could be observed are marked by −. A number of codons which were hitherto never found to be subject of editing, were not investigated experimentally (nd). No potential editing sites are contained in *psaC*, *psaI*, *psaJ*, *psbA*, *psbE*, *psbF*, *psbI*, *psbJ*, *psbL*, *psbM*, *psbT*, *rps15*, *rps19*, *rpl14*, *rpl16*, *rpl36*, *atpH*, *lnfA*, *petE*, *ndhJ*, *ycf6*. *rps16* mRNA which is not encoded by the liverwort plastome was found to be not edited. ((1) Hoch *et al.* (1991); (2) Maier *et al.* (1992a); (3) Maier *et al.* (1992b); (4) Freyer *et al.* (1993); (5) Zeltz *et al.* (1993); (6) Ruf *et al.* (1994); (7) Kössel *et al.* (1993).

whereas some codons were never found to be edited. Interestingly, this preference for certain codon transitions corresponds to the situation observed for plant mitochondria (Araya *et al.*, 1994). The observed bias towards certain codons being subject to editing could be verified by comparing cDNA and genomic sequences of additional maize chloroplast genes. For an estimate of the total number of edited codons we focused our experimental analysis on potential editing sites the codon transition of which had been observed at least once by cDNA sequence analysis derived from maize chloroplast transcripts. As shown in Table 5 we were able to identify seven new editing positions in seven different mRNAs. Inferring that the hitherto not observed codon transitions occur (if at all) with only a very low frequency, a total number of 25, or only a few more, editing sites can be estimated for the maize plastome.

In plant mitochondria, editing is not only confined to protein-coding sequences. Among the more than 400 known plant mitochondrial editing sites, some rare editing sites in rRNAs (Schuster *et al.*, 1991), tRNAs (Maréchal-Drouard *et al.*, 1993) and non-coding sequences (Binder *et al.*, 1992) have been described. Therefore, with respect to other similarities of the two plant organellar editing systems (see above), the existence of a few more editing sites in transcripts, which do not code for peptides, can not be excluded for maize chloroplasts on the basis of the present analysis. However, in view of the overall number of codons in the maize plastome, the estimated total number of codons containing editing sites is very low (0.13%). Thus, in quantitative terms, editing causes only fine tuning of the genetic information defined by a chloroplast DNA sequence. However, in qualitative terms, editing still has to be regarded as an essential prerequisite for chloroplast gene expression. This is evident not only from the few cases in which initiation codons have to be created by editing in order to allow efficient translation of the messages (Hoch *et al.*, 1991; Kudla *et al.* 1992; Neckermann *et al.*, 1994) but also from the mutant phenotype of a transplastomic tobacco plant carrying a non-editable version of the *psbF* transcript (Bock *et al.*, 1994).

Materials and Methods

Cloning

The recombinant pZmc-plasmid clone bank containing *PstI*, *SalI* and *PvuII* fragments of *Zea mays* cp DNA (Fritzsche, 1988) inserted into pBR329 vector were used for cloning of overlapping cp DNA subfragments into pUC19 and pKSII Bluescript vectors. Exonuclease III deletion mutants suitable for complete sequencing of the inserted cp DNA subfragments were obtained according to Henikoff (1987).

Sequencing

Vector-inserted cp DNA was sequenced by the dideoxy chain termination method (Sanger *et al.*, 1977) using fluorescence labelled primers (Igloi & Schiefermayr, 1993). Products of the sequencing reactions were separated by denaturing PAGE and analysed by the automated laser fluorescence system of the EMBL design (Ansorge *et al.*, 1986) and the A.L.F. sequencer (Pharmacia). DNA fragments corresponding to non-cloned gaps of the maize plastome were derived by polymerase chain reaction. Amplification products were sequenced radioactively by a modified chain termination method described by Bachmann *et al.* (1990) or by cycle sequencing using Sequitherm (Biozyme) and fluorescence-labelled primers.

RNA isolation, reverse transcription of RNA, amplification of cDNA and direct sequencing of the amplification products

RNA was isolated according to the method described by Chomcynski & Sacchi (1987). Reverse transcription of RNA primed with hexanucleotide primers in the presence of Moloney murine leukemia virus RNase H-free reverse transcriptase (GIBCO/BRL) was performed according to the manufacturer's instructions. Amplification of cDNA, purification of amplification products and direct sequencing was performed as described (Maier *et al.*, 1992b).

Data analysis

The DNA sequence for a number of maize chloroplast genes had been reported. In order to complete the entire sequence, all known regions were compiled and annotated. In the course of this work several corrections to earlier data were made. This information is provided in EMBL accession no. X86563. Sequence data were compiled and evaluated using the software from Genetics Computer Group (GCG), Madison (Devereux *et al.*, 1984). Database searches were performed with the FASTA and BLITZ algorithms from EMBL, Heidelberg, and the BLAST algorithm available through the blast network service at the National Center for Biotechnology Information (NCBI), USA.

Acknowledgements

We thank Drs Lawrence Bogorad and Masahiro Sugiura for helpful discussions and for encouraging us to complete the sequence analysis of the maize plastome. We thank Elfi Schiefermayr for primer synthesis and assistance with DNA sequencing, Sigrid Krien and Marita Hermann for technical assistance and Dr Elisabeth Fritzsche for constructing the maize chloroplast DNA clone bank. This work was supported by the Deutsche Forschungsgemeinschaft (SFB206) and the Fonds der Chemischen Industrie (H.K.).

References

Ansorge, W., Sproat, B. S., Stegemann, J., Schwager, C. & Zenke, M. (1986). A non-radioactive automated method for DNA sequencing determination. *J. Biochem. Biophys. Methods*. **13**, 315–323.

Araya, A., Bégu, D. & Litvak, S. (1994). RNA editing in plants. *Physiol. Plant.* **91**, 543–550.

Bachmann, B., Lüke, W. & Hunsmann, G. (1990).

Improvement of PCR amplified DNA sequencing with the aid of detergents. *Nucl. Acids Res.* **18**, 1309.

Bedbrook, J. R., Kolodner, R. & Bogorad, L. (1977). *Zea mays* chloroplast ribosomal RNA genes are part of 22,000 base pair inverted repeat. *Cell*, **11**, 739–749.

Binder, S., Marchfelder, A., Brennicke, A. & Wissinger, B. (1992). RNA editing in trans-splicing intron sequences of *nad2* mRNAs in *Oenothera* mitochondria. *J. Biol. Chem.* **267**, 7615–7623.

Bock, R., Hagemann, R., Kössel, H. & Kudla, J. (1993). Tissue and developmental specific modulation of RNA editing of the *psbF* and *psbL* transcript from spinach plastids-evidence for a new regulatory mechanism? *Mol. Gen. Genet.* **240**, 238–244.

Bock, R., Kössel, H. & Maliga, P. (1994). Introduction of a heterologous editing site into the tobacco plastid genome: the lack of RNA editing leads to a mutant phenotype. *EMBO J.* **13**, 4623–4628.

Bubunenko, M. G., Schmidt, J. & Subramanian, A. R. (1994). Protein substitution in chloroplast ribosome evolution. *J. Mol. Biol.* **240**, 28–40.

Chomczynski, P. & Sacchi, N. (1987). Single step method of RNA isolation by guanidinium-thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156–159.

Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 387–395.

Doebley, J. & Stec, A. (1991). Genetic analysis of the morphological differences between maize and teosinte. *Genetics*. **129**, 285–295.

Doebley, J., Renfroe, W. & Blanton, A. (1987). Restriction site variation in the *Zea* chloroplast genome. *Genetics*, **117**, 139–147.

Egli, M. A., Gengenbach, B. G., Gronwald, J. W., Somers, D. A. & Wyse, D. L. (1993). Characterization of maize acetyl-coenzyme A carboxylase. *Plant Physiol.* **101**, 499–506.

Freyer, R., Hoch, B., Neckermann, K., Maier, R. M. & Kössel, H. (1993). RNA editing in maize chloroplasts is a processing step independent of splicing and cleavage to monocostronic mRNAs. *Plant J.* **4**, 621–629.

Fritzsche, E. (1988). Untersuchungen zur molekularen Feinstruktur plastidärer DNA: Alkalilabile Stellen, replikative Intermediate. Kaliumpermanganat als Agens zur Identifizierung von 5'-Methylcytosinpositionen. Doctoral thesis, Fakultät für Biologie der Universität Freiburg.

Fukuzawa, H., Kohchi, T., Shirai, H., Ohyama, K., Umesono, K., Inokuchi, H. & Ozeki, H. (1986). Coding sequences for chloroplast ribosomal protein S12 from the liverwort, *Marchantia polymorpha*, are separated far apart on the different DNA strands. *FEBS Letters*, **198**, 11–14.

Giese, K., Subramanian, A. R., Larrinua, I. M. & Bogorad, L. (1987). Nucleotide sequence, promoter analysis, and linkage mapping of the unusually organized operon encoding ribosomal proteins S7 and S12 in maize chloroplast. *J. Biol. Chem.* **262**, 15251–15255.

Gray, M. W. (1991). Origin and evolution of plastid genomes and genes. In *Cell Culture and Somatic Cell Genetics of Plants*, vol. 7A, *The Molecular Biology of Plastids* (Vasil, I. K. & Bogorad, L., eds.), pp. 5–53. Academic Press, San Diego, CA.

Hallick, R. B., Hong, L., Drager, R. G., Favreau, M. R., Monfort, A., Orsat, B., Spielmann, A. & Stutz, E. (1993). Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucl. Acids Res.* **21**, 3537–3544.

Henikoff, S. (1987). Unidirectional digestion with exonuclease III in DNA sequence analysis. *Methods Enzymol.* **155**, 156–165.

Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Shun, C.-R., Meng, B.-Y., Li, Y.-Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K. & Sugiura, M. (1989). The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.* **217**, 185–194.

Hirose, T., Wakasugi, T., Sugiura, M. & Kössel, H. (1994). RNA editing of tobacco *petB* mRNAs occurs both in chloroplasts and non-photosynthetic proplastids. *Plant Mol. Biol.* **26**, 509–513.

Hoch, B., Maier, R. M., Appel, K., Igloi, G. L. & Kössel, H. (1991). Editing of a chloroplast mRNA by creation of an initiation codon. *Nature*, **353**, 178–180.

Igloi, G. L. & Schiefermayr, E. (1993). Enzymatic addition of fluorescein- or biotin-riboUTP to oligonucleotides results in primers suitable for DNA sequencing and PCR. *Biotechniques*, **15**, 486–497.

Katayama, H. & Ogihara, Y. (1993). Structural alterations of the chloroplast genome found in grasses are not common in monocots. *Curr. Genet.* **23**, 160–165.

Koch, W., Edwards, K. & Kössel, H. (1981). Sequencing of the 16 S-23 S spacer in a ribosomal RNA operon of *Zea mays* reveals two split tRNA genes. *Cell*, **25**, 203–213.

Konishi, T. & Sasaki, Y. (1994). Compartmentalization of two forms of acetyl-CoA caboxylase in plants and the origin of their tolerance.towards herbicides. *Proc. Natl Acad. Sci. USA*, **91**, 3598–3601.

Kössel, H., Hoch, B., Maier, R. M., Igloi, G. L., Kudla, J., Zeltz, P., Freyer, R., Neckermann, K. & Ruf, S. (1993). RNA editing in chloroplasts of higher plants. In *Plant Mitochondria* (Brennicke, A. & Kück, U., eds), pp. 55–64. VCH Publishers, Weinheim.

Kudla, J., Igloi, G. L., Metzlaff, M., Hagemann, R. & Kössel, H. (1992). RNA editing in tobacco chloroplasts leads to the formation of a translatable *psbL* mRNA by a C to U substitution within the initiation codon. *EMBO J.* **11**, 1099–1103.

Kuntz, M., Camara, B., Weil, J.-H. & Schantz, R. (1992). The *psbL* gene from bell pepper (*Capsicum annuum*): plastid RNA editing also occurs in non-photosynthetic chromoplasts. *Plant Mol. Biol.* **20**, 1185–1188.

Larrinua, I. M., Muskavitch, K. M. T., Gubbins, E. J. & Bogorad, L. (1983). A detailed restriction endonuclease site map of the *Zea mays* plastid genome. *Plant Mol. Biol.* **2**, 129–140.

Maier, R. M., Döry, I., Igloi, G. L. & Kössel, H. (1990). The *ndh*H genes of graminean plastomes are linked with the junctions between small single copy and inverted repeat regions. *Curr. Genet.* **18**, 245–250.

Maier, R. M., Hoch, B., Zeltz, P. & Kössel, H. (1992a). Internal editing of the maize *ndh*A transcript restores codons for conserved amino acids. *Plant Cell*, **4**, 609–616.

Maier, R. M., Neckermann, K., Hoch, B., Akhmedov, N. B. & Kössel, H. (1992b). Identification of editing positions in the *ndhB* transcript from maize chloroplasts reveals sequence similarities between editing sites of chloroplasts and plant mitochondria. *Nucl. Acids Res.* **20**, 6189–6194.

Maréchal-Drouard, L., Kuntz, M. & Weil, J.H. (1991). tRNAs and tRNA genes of plastids. In *Cell Culture and Somatic Cell Genetics of Plants*, vol 7A, *The Molecular*

Biology of Plastids (Vasil, I. K. & Bogorad, L., eds), pp. 169–189, Academic Press, San Diego, CA.

Maréchal-Drouard, L., Ramamonjisoa, D., Cosset, A., Weil, J. H. & Dietrich, A. (1993). Editing corrects mispairing in the acceptor stem of bean and potato mitochondrial phenylalanine transfer RNAs. *Nucl. Acids Res.* 21, 4909–4914.

McIntosh, L., Poulsen, C. & Bogorad, L. (1980). Chloroplast gene sequence for the large subunit of ribulose bisphosphate carboxylase of maize. *Nature*, 288, 556–560.

Michel, F. & Dujon, B. (1983). Conservation of RNA secondary structure in two intron families including mitochondrial-, chloroplast- and nuclear encoded members. *EMBO J.* 2, 33–38.

Moon, E. & Wu, R. (1988). Organization and nucleotide sequence of genes at both junctions between the two inverted repeats and the large single-copy region in the rice chloroplast genome. *Gene*, 70, 1–12.

Morton, B. R. & Clegg, M. T. (1993). A chloroplast DNA mutational hotspot and gene conversion in a noncoding region near *rbc*L in the grass family (Poaceae). *Curr. Genet.* 24, 357–365.

Neckermann, K., Zeltz, P., Igloi, G. L., Kössel, H. & Maier, R. M. (1994). The role of RNA editing in conservation of start codons in chloroplast genomes. *Gene*, 146, 177–182.

Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T. & Ohyama, K. (1992). Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. *J. Mol. Biol.* 223, 1–7.

Ogihara, Y., Terachi, T. & Sasakuma, T. (1991). Molecular analysis of the hot spot region related to length mutations in wheat chloroplast DNAs. I. Nucleotide divergence of genes and intergenic spacer regions located in the hot spot region. *Genetics*, 129, 873–884.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S.-T., Inokuchi, H. & Ozeki, H. (1986). Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA, *Nature*, 322, 572–574.

Palmer, J. D. (1983). Chloroplast DNA exists in two orientations. *Nature*, 301, 92–93.

Palmer, J. D. (1991). Plastid chromosomes: structure and evolution. In *Cell Culture and Somatic Cell Genetics of Plants*, vol 7A, *The Molecular Biology of Plastids* (Vasil, I. K. and Bogorad, L., eds), pp. 5–53, Academic Press, San Diego, CA.

Palmer, J. D. & Thompson, W. E. (1982). Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell*, 29, 537–550.

Prombona, A. & Subramanian, A. R. (1989). A new rearrangement of angiosperm chloroplast DNA in rye (*Secale cereale*) involving translocation and duplication of the ribosomal *rps*15 gene. *J. Biol. Chem.* 264, 19060–19065.

Richards, C. M., Hardison, R. C. & Boyer, C. D. (1994). Expression of the large plastid gene, ORF2280, in tomato fruits and flowers. *Curr. Genet.* 26, 494–496.

Rodermel, S. & Bogorad, L. (1987). Genetic map of the *Zea mays* plastid chromosome. *Maize Genet. Corp. News Letters*, 62, 148.

Rodermel, S. & Bogorad, L. (1988). Genetic map of the *Zea mays* plastid chromosome. *Maize Genet. Corp. News Letters*, 63, 155.

Rodermel, S. & Bogorad, L. (1989). Genetic map of the *Zea mays* plastid chromosome. *Maize Genet. Corp. News Letters*, 64, 164.

Ruf, S., Zeltz, P. & Kössel, H. (1994). Complete RNA editing of unspliced and dicistronic transcripts of the intron-containing reading frame IRF170 from maize chloroplasts. *Proc. Natl Acad. Sci. USA*, 91, 2295–2299.

Sanger, F., Nicklen, S. & Coulsen, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA*, 74, 5463–5467.

Sasaki, Y., Hakamada, K., Suama, Y., Nagano, Y., Furusawa, I. & Matsuno, R. (1993). Chloroplast-encoded protein as a subunit of acetyl-CoA carboxylase in pea plants. *J. Biol. Chem.* 268, 25118–25123.

Schuster, W., Ternes, R., Knoop, V., Wissinger, B. & Brennicke, A. (1991). Distribution of RNA editing sites in *Oenothera* mitochonrial mRNAs and rRNAs. *Curr. Genet.* 20, 3974–4004.

Schwartz, Z. & Kössel, H. (1979). Sequencing of the 3'-terminal region of a 16 S rRNA gene from *Zea mays* chloroplast reveals homology with *E. coli* 16 S rRNA. *Nature*, 279, 520–522.

Schwartz, Z. & Kössel, H. (1980). The primary structure of 16 S rDNA from *Zea mays* chloroplast is homologue to *E. coli* rRNA. *Nature*, 283, 739–742.

Shimada, H. & Sugiura, M. (1991). Fine structural features of the chloroplast genome: Comparison of the sequenced chloroplast genomes. *Nucl. Acids Res.* 19, 983–995.

Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B. Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Thodoh, N., Shimada, H. & Sugiura, M. (1986). The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. *EMBO J.* 5, 2043–2049.

Sugita, M., Kato, A., Shimada, H. & Sugiura, M. (1984). Sequence analysis of the junctions between a large inverted repeat and single-copy regions in tobacco chloroplast DNA. *Mol. Gen. Genet.* 194, 200–205.

Sugiura, M. (1992). The chloroplast genome. *Plant Mol. Biol.* 19, 149–168.

Torazawa, K., Hayashida, N., Obokata, J., Shinozaki, K. & Sugiura, M. (1986). The 5' part of the gene for ribosomal protein S12 is located 30 kbp downstream from its 3' part in tobacco chloroplast genome. *Nucl. Acids Res.* 14, 3143.

Vera, A. & Sugiura, M. (1994). A novel RNA gene in the tobacco plastid genome: its possible role in the maturation of 16 S rRNA. *EMBO J.* 13, 2211–2217.

Wakasugi, T., Tsudzuki, J., Ito, S., Nakashima, K., Tsudzuki, T. & Sugiura, M. (1994). Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proc. Natl Acad. Sci. USA*, 91, 9794–9798.

Wolfe, K. H. (1994). Similarity between putative ATP-binding sites in land plant plastid ORF2280 proteins and the FtsH/CDC48 family of ATPases. *Curr. Genet.* 25, 379–383.

Wolfe, K. H., Li, W.-H. & Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl Acad. Sci. USA*, 84, 9054–9058.

Wolfe, K. H., Morden, C. W. & Palmer, J. D. (1992). Function

and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proc. Natl Acad. Sci. USA*, 89, 10648–10652.

Zeltz, P., Hess, W. R., Neckermann, K., Börner, T. & Kössel, H. (1993). Editing of the chloroplast *rpoB* transcript is independent of chloroplast translation and shows different patterns in barley and maize. *EMBO J.* 12, 4291–4296.

Edited by N. H. Chua

(Received 14 March 1995; accepted 9 June 1995)

Transgene Expression in chloroplasts

Agronomic traits	Gene	Promoter	5'/3' Regulatory elements	Reference
Insect resistance	Cry1A(c)	Prrn	*rbc*L / T*rps*16	Mc Bride et al 1995
Herbicide resistance	CP4 (petunia)	Prrn	ggagg / T*psb*A	Daniell et al 1998
Insect resistance	Cry2Aa2	Prrn	ggagg (native) / T*psb*A	Kota et al 1999
Herbicide resistance	CP4 (bacterial or synthetic)	Prrn	*rbc*L or T7 gene 10 / T*rps*16	Ye at al 2001
Insect resistance	Cry2Aa2 operon	Prrn	Native 5'UTRs / T*psb*A	DeCosa et al 2001
Disease resistance	MSI-99	Prrn	ggagg / T*psb*A	DeGray et al 2001
Salt and drought tolerance	*tps*	Prrn	ggagg / T*psb*A	Lee et al 2003
Phytoremediation	*mer*A[a]/*mer*B[b]	Prrn	ggagg[a, b] / T*psb*A	Ruiz et al 2003

Biopharmaceutical proteins	Gene	Promoter	5'/ 3' regulatory elements	% tsp expression	Reference
Protein based polymer	EG121	Prrn	T7gene10 / T*psb*A	Not tested	Guda et al 2000
Human somatotropin	*hST*	Prrn[a], P*psb*A[b]	T7gene10[a] or *psb*A[b] / T*rps*16	7.0 %[a] and 1.0%[b]	Staub et al 2000
Cholera toxin	*ctx*B	Prrn	ggagg / T*psb*A	4%	Daniell et al 2002
Tetanus toxin	*Tet*C (bacterial and synthetic)	Prrn	T7 gene 10[a], *atp*B[b] / T*rbc*L	25%[a], 10%[b]	Tregoning et al 2003
Human Serum Albumin	*hsa*	Prrn[a], P*psb*A[b]	ggagg[a], *psb*A[b] / T*psb*A	0.02%[a], 11.1%[b]	Fernandez-San Milan et al 2003
Interferon alpha 5	INFα5	Prrn	P*psb*A/T*psb*A	ND	Torres
Interferon alpha 2B	INFα2B	Prrn	P*psb*A/T*psb*A	19%	Falconer
Interferon gamma	*ifn-g*	P*psb*A	P*psb*A/T*psb*A	6%	Leelavathi and Reddy, 2003
Monoclonal antibodies		Prrn	ggagg / T*psb*A	ND	Daniell et al (photosynthesis)
Insulin like growth factor	*Igf*-1	Prrn	P*psb*A/T*psb*A	33%	Ruiz G
Anthrax protective antigen	*Pag*	Prrn	P*psb*A/T*psb*A	4-5%	Watson
Plague vaccine	*Ca*F1~*Lcr*V	Prrn	P*psb*A/T*psb*A	4.6 %	Singleton